

All the President's yes-men?

George W. Bush's administration stands accused of biasing the process by which the US government obtains scientific advice. There is a strong case to answer, but the situation is not as unusual as it might at first seem.

The relationship between science and politics is never perfect, but critics charge that the current US administration has so politicized the provision of scientific advice that it could permanently undermine public trust.

Just last week, a storm of protest greeted the announcement that Jerry Thacker, an HIV-positive Christian activist who has referred to AIDS as a "gay plague", would be appointed to the Presidential Advisory Commission on HIV and AIDS. Three months before, a committee advising the Department of Health and Human Services on protecting volunteers in clinical trials was asked to consider whether embryos should be included within its remit (see *Nature* **420**, 3–4; 2002) — a move that critics saw as part of a wider anti-abortion agenda.

The controversy extends to committees that review grant applications. Potential appointees to the panel advising the National Institute for Occupational Safety and Health, for instance, were asked their views on office safety standards — just one example, critics allege, of political considerations impinging on appointments that should depend on scientific merit. The fear is that scientists will refuse to serve on panels that are seen as rubber stamps for administration policies, undermining the quality of the advice given to government agencies and eroding public trust.

Some of the recent developments are disturbing. If the committee on human research subjects gets bogged down discussing abortion politics instead of how to protect patients in clinical trials, lives could be put at risk. Members of committees reviewing grant applications should be selected for their scientific expertise, not their political views.

But successive US administrations, both Republican and Demo-

crat, have packed advisory committees with scientists and other experts who share their political outlook. This only becomes a major issue for the scientific community when the views in question jar with its majority opinion, or the politicization is blatant.

Those with long memories say that the present outcry is reminiscent of the furore inspired by Ronald Reagan's administration in the early 1980s, when it tried similar tactics with committees advising the Environmental Protection Agency — then seen as a thorn in the side of the administration's pro-business policies. This sorry episode alienated environmental scientists, but thankfully the administration eventually backed off and most of the damage was repaired.

There is some comfort to be gained from the checks and balances inherent to the system. The degree of transparency in the formulation of science-led policy in the United States has few parallels in the rest of the world. It is rare indeed for the public to be able to influence government decisions about who sits on the panels and what they discuss. And so far, public input seems to be having a positive effect — last week's storm led Thacker to withdraw from the HIV panel.

This does not mean that the critics should relax. They should look back at the actions of previous administrations to determine the extent to which the current moves represent a departure from accepted practice. The National Academies' Committee on Science, Engineering, and Public Policy is set to take up these questions at its next meeting on 19 February, providing a welcome and timely forum.

Scientists should fight undue attempts by the Bush administration to politicize the advisory process, and extend the same scrutiny to future administrations, whatever their political persuasion. ■

Too quiet on the Eastern front

Change is needed before the nations poised to join the European Union can reap the full scientific benefits of membership.

In May next year, the European Union (EU) will gain 75 million citizens at a single stroke, with the accession of ten new member countries, most of them in Central and Eastern Europe. Among those brought into Brussels' embrace will be some 250,000 researchers — a 15% rise in the EU's scientific workforce. But they won't come with bulging wallets: the eastwards expansion will increase the union's total spending on research by only 3.5%.

Poland, home to more than half of the new EU citizens, provides a prominent example of the scientific weaknesses of the unions' budding members (see page 471). Although the Polish government is stressing that the country's future should not be one of cheap labour, it has done little to nurture a knowledge-based economy. Research spending stagnates at a shameful 0.34% of public expenditure.

But money is only part of the problem. Poland, like many of its neighbours, has too many scientists who aren't internationally competitive, and who cling to the certainties of the socialist era. Unfortunately, many of them sit in senior positions and intend to stay there, which sends genuine achievers to seek opportunities abroad.

Streamlining the overstuffed research system, and letting in the fresh air of competition, is an urgent necessity. The talent exists, and

several research groups are finding niches where Poland can successfully contribute to state-of-the-art research.

Much now depends on another ray of hope: science minister Michał Kleiber. A respected physicist and engineer who joined the government last year, Kleiber has already convinced his colleagues to elevate the State Committee for Scientific Research to a fully fledged ministry. That should give him more clout to push through reforms, but the task remains devilishly hard.

The old guard is bound to fight efforts to concentrate funding into groups that can compete on the European stage, and to close labs that just aren't good enough. But Kleiber's political peers should back him through any short-term unpopularity. Poland will change under EU membership, and some of those changes will hurt. If its research base can be successfully reformed, there is much to be gained in terms of collaboration with Western European scientists — not to mention research grants from Brussels.

The same issues face many of other nations queuing up to join the EU. It would be satisfying to see the largest of the new countries leading the way towards the promised land of European scientific integration, rather than dragging its feet. ■





Little weeds
Experiment casts doubt on effects of transgene flow
p462



Academic ire
Universities angered by Taiwan institute's plan to award degrees
p464



Under fire
Hail of protest greets Italy's scheme for research reforms
p465



Deserted post
Columbia University may pull plug on Biosphere 2
p466

Gates ploughs millions into plan for assault on killer diseases

Declan Butler, Paris

Microsoft founder Bill Gates is to provide US\$200 million for basic biological research into the treatment and control of killer diseases such as malaria and tuberculosis.

The initiative, a partnership between the Bill & Melinda Gates Foundation in Seattle, Washington, and the US National Institutes of Health (NIH), will establish several large consortia to research diseases that kill millions of people in poor countries, but which are often neglected by biomedical research programmes in rich countries.

Gates announced the Grand Challenges in Global Health initiative on 26 January at the World Economic Forum's annual meeting in Davos, Switzerland. It will support about ten large collaborative projects, costing some \$20 million each. Likely targets include work to exploit the recently sequenced genome of the human malaria parasite *Plasmodium falciparum* (M. J. Gardner *et al. Nature* **419**, 498–511; 2002), and the closer study of natural mechanisms that block the development of full-blown tuberculosis in many people who carry the disease.

The initiative will be administered

through the Foundation for the National Institutes of Health, a charity set up in 1996 to help the NIH to collaborate with outside philanthropies and industry. Its scientific programme will be steered by a board chaired by Harold Varmus, former director of the NIH and president of the Memorial Sloan-Kettering Cancer Center in New York. The star-studded board also includes Richard Klausner, former head of the US National Cancer Institute and now executive director of the Gates Foundation's global health programme; Anthony Fauci, director of the US National Institute of Allergy and Infectious Diseases, and Elias Zerhouni, the current NIH director.

Zerhouni says that the initiative will support "high-risk, high-impact science" and complement the NIH's main funding vehicle of investigator-initiated grants. The idea, he says, is to attract top scientists from around the world, including more from the NIH, to global health.

"There are not enough top-flight scientists committed to global health at the fundamental science level," Zerhouni says. The foundation's freedom to fund overseas ►



Bill Gates wants to redress the funding imbalance against research into poor countries' diseases.

M. EULER/AP

UK cell biologist takes pole position at Rockefeller

Erika Check, Washington

Paul Nurse, the head of Britain's largest cancer-research organization, is crossing the Atlantic to take charge of Rockefeller University in New York.

Rockefeller announced the appointment of Nurse, a cell biologist, as its new president on 24 January. Nurse plans to take up the position full-time in November, and says that he intends to focus on developing the university's ability to do interdisciplinary and translational research.

"Rockefeller has tremendous strengths, particularly in cell biology and aspects of organismal biology and neurobiology," Nurse says. "And it has a high-quality research hospital, which allows you to explore how all this fantastic biology can



Paul Nurse: all set for move to New York.

lead to new therapies."

The Manhattan-based institution is unique among major US research universities in having no undergraduate students or departments, but instead operating 70 fairly autonomous laboratories.

Previous presidents of Rockefeller have had a bumpy ride. The last president, Arnold Levine, left last February after a

strange public encounter with a graduate student (see *Nature* **415**, 721; 2002). And back in 1991, David Baltimore had to resign over his handling of a case of alleged

scientific misconduct involving a co-worker.

Nurse, who won a Nobel prize in 2001 for his work on the cell cycle, led one of Britain's main cancer charities, the Imperial Cancer Research Fund, into its merger last year with the Cancer Research Campaign to create Cancer Research UK, which he has since run.

"Paul Nurse has the stature befitting a Rockefeller president," says David Ho, an AIDS researcher at Rockefeller's Aaron Diamond AIDS Research Center. "He's had so much background experience leading the cancer effort in the UK — and he has tremendous scientific credibility."

Nurse has never worked full-time in the United States. "Many researchers go to work in other countries early in their career," he says. "I'm just doing it at the end of mine." ■

is also “a potential means to enlarge our research mission internationally”, he says.

Jacqueline Fuller, a spokeswoman for the Gates Foundation, says that the initiative will soon embrace other partners to make it truly international in scope. To this end, the foundation is in discussions with the London School of Hygiene and Tropical Medicine, where Gates has already contributed \$40 million to a malaria research centre, and the British research-funding charity the Wellcome Trust.

Gates sees the initiative as a starting point in tackling a huge imbalance whereby less than 10% of the \$70-billion annual world spend on medical research goes on the diseases that cause 90% of illness and death. “There is great potential for science and technology to solve persistent global health challenges, but far greater resources are needed,” he says.

Existing global expenditure on malaria research is estimated to be just \$100 million annually, and the Special Programme for Research and Training in Tropical Diseases run by the United Nations and the World Bank has only \$30 million a year to cover ten major tropical diseases.

Impressive as it, the Gates grant is small beer next to the \$1.5 billion that the World Health Organization has said is needed annually to support a Global Health Research Fund dealing with major killer diseases. But Francis Nkrumah, director of the Noguchi Memorial Institute for Medical Research at the University of Ghana and the only African on the board of the new initiative, says it marks a major step towards that goal. He thinks that it will help to forge a consensus among biologists on the most promising research avenues, and hence attract further private and public research funds. “This is a beacon of hope for Africa,” he says. ■

♦ www.gatesfoundation.org

Transgenic crop trial's gene flow turns weeds into wimps

David Adam, Amsterdam

Could ‘superweeds’ carrying genes from genetically modified crops behave less like Superman and more like Clark Kent, his puny alter ego? The first results from a pioneering field trial in the United States suggest as much — and that the effects of gene flow from transgenic crops may be less aggressive than some environmentalists predict.

The experiment, run by Neal Stewart and his colleagues at the University of Tennessee, Knoxville, would have struggled to win regulatory approval in Europe. In it the team used an oilseed rape crop that had been given a gene from the bacterium *Bacillus thuringiensis* (*Bt*), enabling it to produce a toxin that repels insects. They crossed the genetically modified variety with a wild relative, *Brassica rapa*, then backcrossed the resulting hybrid with the wild plant again, and released the resulting ‘superweed’ into the environment.

The resistant weed’s ability to compete as a pest — and therefore the likelihood of the rogue gene sweeping through wild plant populations — was assessed by its effect on fields of wheat. In other fields, the researchers introduced naturally occurring weeds.

The team found that the transgenic weed was far from dominant, having 20% less effect on wheat yield than the unmodified *B. rapa* weeds. Stewart presented his team’s results, from the first year of trials in North Carolina and Georgia, at a conference on gene flow between plants held in Amsterdam on 21–24 January.

Stewart, who believes that genetically modified crops are currently “over-regulated”, suggests that the modified weeds lose potency because they are disrupted by the genetic load



Early trial results suggest that gene flow from transgenic crops puts weeds at a disadvantage.

of crop genes being carried over with the *Bt* transgene. “Weeds have undergone years of selection that make them very good at what they do,” he says.

Other plant researchers welcomed the results as an important advance. “We need to move on from asking whether gene flow takes place, to investigating what happens when and where it does,” says Brian Johnson, biotechnology adviser to the conservation group English Nature and an expert on transgenic-crop issues.

But some cautioned that the preliminary results are not a green light for the use of such crops in Europe, where their commercial planting is under a *de facto* moratorium. In a separate study announced last year, plant ecologist Allison Snow at Ohio State University in Columbus found that similar *Bt* transgenes can make wild sunflowers produce more seeds — a sign that modified wild populations could prosper and spread in the environment (see *Nature* 419, 655; 2002). ■

Head of Spanish lab network quits over lack of funds

Monica Salomone, Madrid

The president of Spain’s main network of scientific research laboratories has resigned, complaining that the government gave him neither the resources nor the autonomy he needed to do the job.

Theoretical physicist Rolf Tarrach, who ran the CSIC research council for two-and-a-half years, says that he will leave his post within a few weeks, once he has helped the government to select his successor.

Spanish science minister Josep Piqué confirmed the news, saying that Tarrach “had expressed his wish to go back to his research a long time ago”.

At the CSIC, Tarrach was responsible for

120 laboratories, which employ about 2,200 researchers. But he often complained about the lack of funding and human resources. “I could do very little of what I intended” in the position, he says.

In a letter to the science ministry last October, Tarrach had asked to be replaced if certain measures were not taken. He wanted to see salaries for CSIC researchers rise by 7–10% to match those of university researchers. “The best researchers won’t go to the CSIC” unless this happens, he says. “Decisions about the CSIC’s policy that used to be taken by its presidency are now taken at the ministry,” he adds. “This makes it very difficult to run a modern institution.”

Tarrach says that his requests were ignored, so he filed his resignation two weeks ago and will now return to his academic post at the University of Barcelona.

The resignation has yet to be announced publicly. This, says Tarrach, is to avoid “interference” with the storm of criticism currently lashing the Spanish government over its handling of the ecological disaster caused by the sinking of the tanker *Prestige* (see *Nature* 420, 347; 2002).

One possible candidate to replace Tarrach is Emilio Lora-Tamayo, a physicist who is currently vice-president of the CSIC and head of a commission advising the government on its response to the *Prestige* oil spill. ■

US army survey targets Gulf War syndrome

Jonathan Knight, San Francisco

In a push to avoid trouble with mysterious post-conflict syndromes, the Pentagon is stepping up efforts to monitor the health of soldiers before and after their deployment in the Middle East.

Some scientists who have studied 'Gulf War syndrome' — the set of maladies associated with active service in the last Gulf War — applaud the effort. But others doubt whether it will pinpoint the source of health problems, should they occur during any conflict in Iraq.

The effort relies mainly on questionnaires. One, which all soldiers must now fill in before they are deployed and again when they return, asks about the soldier's general health and whether he or she is concerned about "possible exposures" during deployment. Responses are stored in a computer database maintained under a project called the Army Medical Surveillance Activity.

A more extensive questionnaire is being given to participants in the Millennium Cohort, a 21-year health study of some 100,000 service members that began in 1999. The survey, given every three years, tracks



The US army is keen to use experience of the Gulf War to monitor health problems in any Iraq conflict.

the participants' health with questions about sleeplessness, stress symptoms, fitness, and alcohol and tobacco use. It also allows subjects to report their own health concerns.

If a group of veterans with related health problems crops up in either survey, military researchers will use records of troop movements to spot any locations or health hazards

that they might have had in common.

That will be easier today than it was ten years ago, says John Resta, director of health risk management at the US Army Center for Health Promotion and Preventive Medicine in Maryland. Since the Gulf War, the army has stepped up water- and air-quality monitoring, he says, and it now deploys special teams to assess unusual hazards such as radiation and chemical contamination. These tests are linked to daily troop movements in a central database.

But even with all these data, the cause of something akin to Gulf War syndrome would be very hard to track, says Robert Sapolsky, a stress biologist at Stanford University. Too many disparate factors seem to contribute. For example, recent evidence suggests that stress can open the blood-brain barrier, rendering relatively harmless levels of contaminants more dangerous.

Epidemiologist Robert Haley at the University of Texas Southwestern Medical Center in Dallas, who believes some Gulf War veterans suffered brain damage from exposure to nerve gas, says that the surveys may mask such problems. Any brain injuries would almost certainly show up as stress problems in such broadly targeted questionnaires. "They will just say there is lots of post-traumatic stress disorder," he says.

But Simon Wessely, an epidemiologist at King's College London who has led the British effort to understand Gulf War health issues, supports the army's plan to look at a wide array of health issues. Wars invariably produce post-war syndromes, but the symptoms are always different, he says. After the Boer War, many soldiers complained of rheumatism, whereas stomach problems and fatigue prevailed after the Second World War (E. Jones *et al. Br. Med. J.* 324, 321–324; 2002). So it makes sense to cast a wider net. "You can't say what the injuries will be, only that there will be injuries," Wessely says.

TB medic wins global health post

Declan Butler, Paris

Jong Wook Lee, a South Korean physician, is set to succeed Gro Harlem Brundtland as director-general of the World Health Organization (WHO).

Lee, who has worked for the agency for 19 years, emerged victorious after a secret ballot by the agency's 32-member executive board on 28 January. It was a close finish with Peter Piot, a former AIDS researcher who now heads the Joint UN Programme on HIV/AIDS (UNAIDS). Both men were neck-and-neck after the second round of voting, with Lee clinching the win in a tiebreaker vote. Pascoal Mocumbi, Mozambique's prime minister, who had been tipped as a favourite for the post (see *Nature* 421, 302; 2003), was eliminated in the first round.

Lee has served as head of the WHO's vaccine and immunization programme, and since 2000 has fronted the agency's Stop TB programme. This widely commended initiative involves a consortium of 250 partners including WHO member states, donors, industry and non-governmental organizations.

In his election manifesto, Lee said that targeted investment in particular diseases was not sufficient to make a dent in the global health burden. He promised to press for "substantial investments in health services".



Jong Wook Lee: ready to head the World Health Organization.

Lee also said that he would continue Brundtland's efforts to restructure the WHO, which at the time she took over was widely criticized for being an ineffective, excessively politicized bureaucracy.

Brundtland succeeded in putting the WHO and global health issues such as malaria and TB on the international agenda. But she was less successful in her reforms to streamline the agency, and many staff complain of an aloof management culture.

Lee has pledged to boost the low morale among WHO staff, and launch employment programmes to attract the best recruits. He also plans to decentralize the management of WHO programmes to relevant regions, for example by shifting the leprosy programme to India, and traditional medicine to its western Pacific regional office.

The nomination must now be approved by the WHO's 192 member states at the 56th World Health Assembly in Geneva in May, with the new director-general beginning his five-year term on 21 July.

Taiwanese institute causes upset by degrees

David Cyranoski, Taipei

Taiwan's premier research institute is facing dissent from local universities over its planned involvement in the country's graduate education programme. The universities are concerned that Academia Sinica, which is pushing for a change in the law to allow it to admit graduate students, is muscling in on their territory.

Academia Sinica, which comprises 26 institutes and almost 1,000 researchers in a wide range of disciplines, is aiming to reduce Taiwan's reliance on the United States and Europe to train its young scientists. But Taiwan's universities — currently the only bodies that are able to award degrees — argue that they are best placed to offer balanced postgraduate training.

Some researchers think the universities' real worry is that Academia Sinica's prestige, together with its relatively generous stipends, could eventually deprive them of the country's best young scientific talent. In an effort to assuage this worry, the research institute is pledging to award three-quarters of its doctorate posts to students from outside Taiwan.

Yuan Tseh Lee, president of Academia Sinica, says he wants to create a world-class, international graduate school at the institute. "We want train students to international standards so they can go abroad — and also bring international talent here," he says.



Academia Sinica wants to share the right of Taiwan's universities to award degrees.

Until now, Academia Sinica has run a collaborative programme that admits graduate students to one of Taiwan's premier universities, with which it shares responsibility for their training.

Last autumn, two such programmes took on five students between them. The students each took a first-year course at the academy's Taipei campus, taught in rotation by scientists from the academy and the National Tsing Hua University, based 70 kilometres away in Hsinchu. After the first year, students select an adviser from one of the two institutions. The entire programme is run in English.

This autumn the programme will expand to cover seven subject areas and involve more universities, including the National Taiwan University. The programme's eventual aim is to admit 20 students each year.

But Lee says that the collaborative approach raises administrative problems, and Lawrence Huang, executive secretary of the programme, says that some researchers at Academia Sinica are unhappy that candidates have to sit universities' qualifying examinations, and with the requirement that foreign students study Chinese.

Lee has therefore been lobbying to have the law changed to allow Academia Sinica to award degrees. But the idea has met with hostility from universities. "Lee invited the top university presidents to talk about it," says National Taiwan University president Wei-Jao Chen. "We all said it was a bad idea. Education should include more than research: everyone should have the opportunity for a more general curriculum and interaction with other students." Moreover, he adds, if Academia Sinica runs the programme alone, the universities will lose the chance to work with it and share its resources.

Lee says that such concerns are unfounded, and that collaboration remains important to the institute. Huang agrees: "Our researchers are good at research, but not necessarily at teaching. We have to learn from the universities."

Lee is likely to get his way when the Taiwanese parliament meets this spring, however. Chen is resigned to the idea: "When the law changes we will have to go along with it."

♦ www.sinica.edu.tw/~tigg

ACADEMIA SINICA

Scientists call for Canada to boost polar-shelf funding

David Spurgeon, Montreal

Canadian scientists who study the Arctic say that the country's Polar Continental Shelf Project is in danger of failure, threatening research into the environment and ecology of almost half of the nation's territory.

Forty-two scientists from more than a dozen disciplines sent a letter to Prime Minister Jean Chrétien on 6 January calling for Can\$25 million (US\$16.5 million) to be

added to the present "utterly inadequate" funding for northern science. They claim that Arctic research funding in Canada amounts to only 20 Canadian cents per head of population, compared with about Can\$3 in the United States.

"Our international credibility is at stake and our ability to recruit the next generation of Canadian Arctic scientists is rapidly expiring," the scientists wrote.

Arctic researchers from Canada and abroad rely on the polar-shelf project's resources for air and ground support. It is operated by Canada's natural resources department on an annual budget of some Can\$3.5 million — less than half the amount it received a decade ago — and faces sharp increases in its operating costs.

The 45-year-old project "is the backbone of Canadian Arctic science," says David Hik, an ecologist at the University of Alberta. "We don't understand how, at its current budgetary levels, it can continue to support northern science in any meaningful way."

Greg Henry of the University of British

Columbia, who has conducted one of the world's most northerly climate experiments on Ellesmere Island, says he cannot compare his results with those at other Arctic sites because he can't get enough flight time from the polar-shelf project.

Henry says that Canada, which led world research on the effects of pollution in the Arctic, has lost many of its best researchers through government cuts. The scientists' letter says that the success of international agreements relies on Canadian data that the country is struggling to provide.

In September 2000, a national task force called for more federal funding for Arctic research, proposing the establishment of 24 university chairs, 40 graduate scholarships, 40 postdoctoral fellowships and support for 70 strategic research projects.

"Only about 10% of those recommendations have been implemented," says Hik, adding that the letter to the prime minister is only part of a push by scientists to get the government to follow through on the task force's findings.

PCSP/NATURAL RESOURCES CANADA



Left out in the cold: the Polar Continental Shelf Project's budget has halved over ten years.

Europe seeks single defence research agency

Quirin Schiermeier, Munich

European policy-makers are thinking of setting up a counterpart to the US Defense Advanced Research Projects Agency (DARPA). The US organization enjoys a reputation as a forward-thinking supporter of innovative research.

Greece, which this month took over the rotating presidency of the European Union (EU), is expected to ask the European Commission to draw up a set of options to beef up Europe's military research programmes, and foster multilateral cooperation in fundamental defence-related research.

At a meeting of EU commissioners in Athens earlier this month, Yiannos Papantoniou, the Greek defence minister, said that he favours either an extension of the EU Framework research programme to include military and 'dual use' research, or the creation of an agency similar to DARPA.

DARPA plays an important role in supporting basic, openly published scientific research at US universities, especially in disciplines such as materials, computer science and engineering.

Most of Europe's defence-related research is carried out by national agencies. "This has led to fragmentation and duplication of research," argues Jordi Molas-Gallart, a defence research analyst at the Science and Technology Policy Research unit at the University of Sussex, UK. "There is scope for much more coordination — but there are also entrenched national misgivings and strategic concerns."

The EU's six biggest weapons-buying nations — France, Italy, Spain, Sweden, Germany and Britain — signed a memorandum in 2000 to cooperate on military procurement. But this had no effect on scientific research, says Molas-Gallart.

Last year, a working group of the European Convention, which is to draft a new framework and structures for the EU by 2004, proposed the creation of a European armaments and strategic research agency, which would promote harmonized procurement and research.

But a basic research agency like DARPA, with scientific autonomy and a mission to address generic, high-risk ideas in various disciplines, would take that concept a step further. "A European DARPA, if it avoids EU bureaucracy, could become a clearing house for advanced technologies for military and civilian applications," says Alex Nicoll, assistant director of the London-based International Institute for Strategic Studies.

European officials say that EU research commissioner Philippe Busquin is open to the idea in principle. "It is surely an option, but it is at a very early stage," says a spokesman for the commission.

Nicoll believes there is political will to strengthen the EU's military role, but warns that without a common defence policy, any EU defence-research programme could fail.

The gap in defence spending between the United States and its 17 European allies has been growing. In 2001, the United States spent almost \$30,000 per service person on R&D — twice as much as Britain, more than three times as much as France, and almost ten times as much as Germany, according to the IISS. Basic and applied research related to military needs received \$14 billion in the United States, compared with \$7 billion in France, Britain and Germany combined. ■

► www.weu.int/weag

► www.iiss.org



Italians riled by science reforms

Alison Abbott, Munich

Top Italian researchers have lambasted the research ministry for what they call an unconstitutional effort to grab direct control of the CNR, the country's leading basic-research organization.

The scientists are angry that under a decree made public on 24 January, the research ministry would acquire the power to appoint key figures within the CNR, which would be restructured to focus on applications of research. The researchers also complain that they were not consulted. They have now been given two months to comment on the decree.

Lucio Bianco, president of the CNR, says he will make sure that the legality of the decree is challenged in court if the government refuses to modify its contents. "Autonomy of scientific research institutes,

universities and academies is guaranteed in the Italian constitution," he argues.

The plan would divide the CNR, which runs some 100 institutes, into discipline-related departments whose heads will be appointed by the ministry. Department heads would be responsible for allocating resources and nominating institute directors. The names of the departments, such as 'science and technologies of medicine', or 'science and technologies of advanced production systems', strongly emphasize the applications of research.

Education, university and research minister Letizia Moratti says that her proposed new structure will make the best use of available resources. But Carlo Bernardini, a high-energy physicist from the University of Rome, says it is ill-conceived. "No one argues against making research more productive, but you can't do this by putting in managers to steer basic research," he says. Bernardini is unhappy that the ministry used the consultancy firm Ernst and Young to develop the strategy, "but excluded scientists — even the presidents of basic research organizations".

"Scientists see this as an attack on the autonomy of research," says Rino Falcone, of the CNR Institute of Cognitive Sciences and Technology in Rome, who has organized protests against the decree.

The decree also rearranges some organizations, bringing the INFN, an agency for material sciences that was spun out from the CNR in 1994, back inside it, and transferring the CNR's astrophysics institutes to the Institute of Astronomy, formed by a merger of Italy's 12 observatories in 2000. ■



Research minister Letizia Moratti wants fresh focus for Italy's main basic-research agency.

Russian researchers hail move to bring salaries up to scratch

Moscow More than 50,000 Russian scientists are set to receive a remarkable New Year's bonus this week — an inflation-busting pay rise of up to 500%.

In a determined bid to halt the brain drain of researchers, the Russian Academy of Sciences has announced unprecedented and immediate salary increases for its employees. The move follows a public commitment from the government in November last year to boost support for science.

The 1,250 full and corresponding members of the academy — the élite of Russian researchers — will now receive monthly salaries of 20,000 and 10,000 roubles (US\$630 and \$315), respectively, five times more than they earned before. The salaries of postdocs and assistant professors will rise threefold, to 900 and 1,500 roubles, respectively.

Technicians, laboratory workers and young scientists will also get substantially more money later this year, says Boris Myasoedov, the academy's deputy secretary general for science. "We are very, very happy about this pleasant development," he says.



The Russian Academy of Sciences is offering large salary hikes to combat the country's brain drain.

Stargazers win heated battle over Native American site

Washington Astronomers have finally triumphed in a three-year battle to win approval for the VERITAS observatory on Mount Hopkins in Arizona.

Officials from the Coronado National Forest granted permission on 16 January for the construction of VERITAS (the Very Energetic Radiation Imaging Telescope Array System). The telescope array is designed to study high-energy radiation sources, such as active galactic nuclei and the remnants of supernovae. The site was controversial because it is close to a Native American 'sweat lodge' — a spiritual sauna in which occupants meditate around very hot stones.

Announcing the decision, forest supervisor John McGee said that the impact on the Native American site had been considered, but that "the scientific potential

of the VERITAS justifies these trade-offs".

As a compromise, the access road for the telescope will be located farther from the sweat lodge than originally planned. Pending any new challenges during a 45-day appeal period, and funding from the National Science Foundation, the \$25-million observatory could begin operations in 2005.

Antinori makes a meal of hunger-strike claim

Rome Having seen the Raelian sect draw so much interest over human cloning in recent weeks, fertility doctor Severino Antinori has hit on an unusual method of winning back the spotlight.

In a dramatic statement issued outside Italian government offices on 21 January, Antinori claimed that he was beginning a hunger strike "to the death". The fast is in protest against what he describes as persecution by Italian authorities, who have been attempting to determine whether his Rome-based clinic has carried out banned human-cloning work. Antinori has previously claimed that one of his patients was set to give birth to a cloned baby early this year.

Antinori is now demanding a personal guarantee of Italy's constitutional freedom of research from prime minister Silvio Berlusconi. But one week into his protest, Antinori has not been seen and is declining to comment on either his heroic gesture or his resulting physical health.

British science courses face high fees and low interest

London Science degree courses in England and Wales could become significantly more expensive than other subjects under government proposals for higher education, critics have charged.

A white paper on the future of higher education, published on 22 January, includes plans to allow universities to set their own fees of up to £3,000 (US\$5,000) per year for individual courses. Students currently pay just over £1,000 a year for all subjects.

Science courses cost universities more to provide, and some fear that this could lead to higher price tags — with a resulting slump in interest.

There was some good news for research, however. Annual funding for university teaching and research will rise from the current £5.7 billion to £6.9 billion by 2006. And the government has found an ingenious way to increase research quality at little cost: a new top grade in the Research Assessment Exercise means that the current ratings will be extended to include a 6* grade beyond the existing 5* and the less impressive 1 to 5.

♦ www.dfes.gov.uk

Lab investigators make return after sacking

Washington The University of California has re-hired two independent investigators who were fired last November from the Los Alamos National Laboratory in New Mexico.

The investigators, Steven Doran and Glenn Walp, were dismissed after publicly releasing documents that they claimed showed numerous management failures at the nuclear weapons laboratory, which is operated by the university. Their sacking led to investigations by Congress and the Department of Energy, as well as the resignation of the lab's director and deputy director (see *Nature* 421, 99; 2003).

After a four-hour meeting between the two investigators and Richard Atkinson, the university's president, which both sides described as "candid and constructive", the pair have been re-hired in time for a separate investigation into the university's management of the laboratory.

Cash could dry up for desert ecological experiment

San Francisco Columbia University is reconsidering its funding of Biosphere 2, raising serious doubts over the future of the environmental research centre in the Arizona Desert.

Representatives at the centre say that the university wants to reduce its \$20-million commitment to fund the elaborate ecology experiment until 2010, and may pull the plug altogether. University officials say that they are considering several proposals, but favour forming a consortium with other universities to share the cost. Columbia's plans to hire new researchers at the facility have been dropped, and the masters programme in environmental public policy that it ran there is moving to the university's main campus in New York.

Biosphere 2 has struggled to win scientific credibility since it was opened by Texan billionaire Ed Bass in 1991. Columbia University took over its management in 1996, and has spent \$24 million trying to breathe new life into it (see *Nature* 402, 567; 1999).



End of the road? Prospects are bleak for Biosphere 2 if Columbia University withdraws its funding.

Robots in the deep



Out of the blue: the Autonomous Benthic Explorer (main picture and inset) triumphed when it discovered the small hydrothermal vent known as 'Rose Bud' (right) near the Galápagos Islands.

Having overcome problems with reliability, autonomous submersibles are starting to make their mark on marine science. Tom Clarke follows their progress beneath the waves — without getting his feet wet.

The 'Rose Garden' amazed marine scientists when it was discovered in 1979. Hydrothermal vents — outpourings of volcanically superheated water from the sea bed — had been identified just two years before. But this example surpassed all expectations. Covered in bizarre tubeworms, giant clams and ghost-white crabs — an entire ecosystem powered, ultimately, by sulphurous chemicals in the vent's scalding waters — this deep-sea oasis became the poster child of marine exploration, gracing the cover of *National Geographic* magazine.

So imagine the disappointment last June, when scientists returned to the location, some 370 kilometres northeast of the Galápagos Islands, and could find no trace of the Rose Garden. Had it been engulfed in a seafloor eruption, or were the researchers looking in the wrong place? We might still be in the dark, were it not for a robotic submersible called the Autonomous Benthic Explorer (ABE), developed at the Woods Hole Oceanographic Institution (WHOI) in Massachusetts. By hovering some 40 metres above the seabed, ABE mapped the area while recording water currents and temperatures. It confirmed that the Rose Garden was no

more — but showed that a new vent, nicknamed Rose Bud, had sprung up in its place.

ABE's success in solving the Rose Garden mystery, revealed at the American Geophysical Union's meeting in San Francisco in December, is just one sign that such autonomous underwater vehicles (AUVs) are finally beginning to fulfil their scientific promise. Under development for a decade or more, AUVs are designed to tackle tasks that are too dangerous for crewed vessels, or too time-consuming for remotely operated craft that must be steered from the surface. They can also be sent to ocean regions that would otherwise be inaccessible. Understanding what is going on in the ocean "is about being out there at the right time in the right place", says Jim Bellingham, an engineer at the Monterey Bay Aquarium Research Institute (MBARI) in California. In this lottery of time and place, AUVs hold the promise of shifting the odds in oceanographers' favour.

Over the past three years, AUVs have studied algal blooms off the coasts of the United States, tracked stocks of North Sea herring and surveyed major new oilfields. And in the coming months, they will travel beneath Antarctica's ice shelves, investigate toxic red tides in the Gulf of Mexico, and study

upwellings of nutrient-rich water off the Californian coast. If these pioneering projects prove successful, the oceans of the future could be buzzing with robotic probes — some of them spending months at a time on the move, others lying dormant underwater, waiting to be deployed to study sudden events such as seafloor volcanic eruptions.

Spaced out

Marine scientists like to compare their efforts to study the oceans' vastness to the exploration of space. By this analogy, crewed submersibles are like the Apollo Moon missions: extremely expensive and only able to make fleeting trips. AUVs, on the other hand, could fulfil the role of unmanned space probes, taking payloads of scientific instruments on lengthy voyages to inhospitable locations without risking human life.

At least that has been the theory. In practice, however, AUVs have long had a reputation for being unreliable and unpredictable. ABE's first mission, in 1994, was nearly its last, when the device that jettisons its ballast failed. Unable to ascend from the seabed, it had to be rescued by a crewed submersible. Autosub, an AUV developed at Britain's



D. FORNARI/WHOI

WHOI

Southampton Oceanography Centre, suffered similar ignominy in 2000, when it got stuck beneath an overhanging cliff on the sea floor off the Strait of Sicily. With no reverse gear, it was unable to surface, as it is programmed to do in emergencies.

But advances in distributed network architecture, which coordinates the different computers that govern the aquatic robots' components, together with more sophisticated navigation tools and the growing experience of the crafts' handlers, are now turning AUVs' reputation around. "Not only has the technology matured, it's done so very rapidly," says Gwyn Griffiths, head of the Ocean Technology Division at the Southampton centre.

Paths of discovery

With the worst of their reliability problems hopefully behind them, AUVs have begun to prove themselves. ABE's mission to the Galápagos Rift, home of the Rose Garden, illustrates the potential. Sluggish but highly manoeuvrable, ABE can be programmed to follow a predetermined path, dropped over the side of a research vessel, and left to carry out its task. This allowed the scientists on last June's cruise to leave ABE to investigate the Rose Garden mystery, while they got on with other projects and caught up on some sleep.

Some 2.5 kilometres below the surface, ABE moved back and forth in the gloom in a systematic pattern — "mowing the lawn", as WHOI engineer Dana Yoerger puts it. Using sonar, the craft produced a topographical map of the seabed with 30 times the resolution attainable from a surface ship. "It's like someone switched the lights on," says Yoerger. For those exploring the Galápagos Rift, this illumination is crucial. Anna-Louise Reysenbach, a microbial ecologist at Portland State University in Oregon, who visited the rift in the crewed submersible *Alvin* shortly after ABE's survey, says it was like "hiking with a good map". Previously, she'd have been lost in the wilderness.

The discovery of Rose Bud — the smallest and possibly the youngest hydrothermal vent ever found — is another feather in ABE's cap. The craft's temperature sensor detected a discrete, rising plume of water that was 0.02°C warmer than the surrounding ocean. This anomaly, too small for *Alvin* to have detected, led researchers to Rose Bud. This tiny vent is just a year or so old, but is already developing its own population of deep-sea creatures, as subsequent visits in *Alvin* have shown.

Whereas ABE is designed to explore the ocean's depths, other AUVs have been built to range far and wide through waters nearer to the surface. The torpedo-shaped Autosub is one such craft. In the past few years, it has found its sea legs, by showing that herring aren't scared away by the engine noise of ships conducting sonar surveys of fish stocks¹, and surveying the whereabouts of schools of krill under Antarctic sea ice².

Next month, Autosub will embark on

perhaps the most challenging AUV mission so far. Led by Adrian Jenkins of the British Antarctic Survey in Cambridge, researchers are sending the 7-metre-long vessel beneath the ice shelf at the base of the Pine Island Glacier in Antarctica, where it floats on the sea (see map, below). Such ice shelves, which form where polar glaciers meet the sea, are central to the recycling of fresh water in the form of snow and ice back into the oceans. About 1,000 cubic kilometres of fresh water are thought to melt into the ocean from the underside of Antarctic ice shelves each year, and measurements of ice volumes made at the surface and of the speed of ice-shelf glacier movement suggest that this process may be speeding up³. Yet no one has studied what's happening beneath the ice directly. "It's a total black box," says Jenkins. "We just don't know what's driving the changes there."

Venturing under the Pine Island shelf, which measures some 70 by 30 km, Autosub will use temperature and salinity sensors to measure the rate of melting along its underside. It will deploy an acoustic doppler profiler to detect ocean currents, and sonar to map the topography of both the sea bed below and the ice above. A sampler on board will also return water from beneath the shelf for lab-based analysis. With a range of 500 km and no need to communicate with the surface, Autosub is one of the few devices that are capable of taking these measurements in such a hazardous environment. "It opens a whole new window onto processes within polar oceans," says Jenkins.

If successful, the Pine Island mission will be followed next year by a more ambitious excursion under the Ronne Ice Shelf — which is the size of France and floats on the Weddell Sea. Autosub is fitted with a state-of-the-art gyroscopic compass, and will also

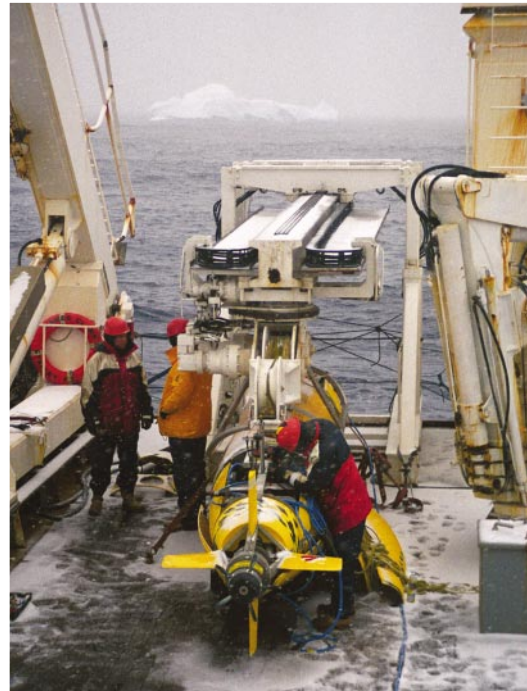
Chilly challenge: Autosub (right) is set to study freshwater melting under the Antarctic ice at Pine Island and, if successful, the Ronne Ice Shelf.



carry an emergency acoustic beacon to signal for help if it gets into trouble. But if anything goes seriously wrong any distance from the edge of an ice shelf, Autosub's handlers accept that they will never see the craft again. "With everything you put in the sea, you have to be prepared to lose it," says Griffiths. At least, he muses, the £850,000 (US\$1.37-million) submersible "would have gone down in the service of science".

Other researchers are working with smaller, cheaper AUVs, where the stakes are somewhat lower. Since the late 1990s, researchers at the Coastal Ocean Observation Lab at Rutgers University in New Brunswick, New Jersey, have been using 1.5-metre Remote Environmental Monitoring Units (REMUS), originally developed at WHOI, to study coastal processes. These AUVs look like miniature versions of Autosub, and operate for only a few hours, as opposed to days, at a time. Even over such short periods, however, they can be deployed from near the shore to study the movement of sediment, the role of waves in oxygenating water, or the effects on the coastal environment of algal blooms and discharges from rivers. "You can put them in the back of a truck and drive them to where they're needed," says Scott Glenn, an oceanographer at Rutgers.

Glenn's team is now starting to deploy



SOUTHAMPTON OCEANOGRAPHY CENTRE

REMUS in combination with graceful long-duration devices called Slocum gliders, to study the formation of toxic 'red tides', caused by algal blooms, in the Gulf of Mexico off Florida. The gliders, around 2 metres long, operate on much the same principle as their airborne counterparts. Their 'wings' allow them to cover long distances while gradually sinking through the water, and a rudder enables them to stay on course. On reaching the bottom of a predetermined trajectory, the gliders then use a battery-powered pump to change their buoyancy and float straight to the surface. There they use a global-positioning-system device to establish their location, adjust their rudder accordingly and glide downwards again. Because they don't rely on a power-sapping motor, the gliders can sawtooth their way back and forth across an area of interest for days or even weeks at a time.

Glenn hopes that the combination of gliders and REMUS will help to explain how the mysterious red tides form, perhaps revealing whether they are natural phenomena or a consequence of human activities such as agricultural run-off or watershed management. The gliders will provide baseline information on the presence of algae and currents, and on water temperature and salinity, while REMUS will be deployed to obtain more detailed measurements from the most interesting parts of an algal bloom.

At MBARI, Bellingham has also been working with combinations of gliders and powered AUVs — in his case, mid-sized craft called Dorados. This summer, he will contribute to a US Navy-funded study of upwelling in Monterey Bay. As well as supporting a huge array of marine life, upwelling zones are ideally suited to studying mixing of ocean water of different temperatures and salinities. The project, which involves 20 institutions, will deploy sensors on aircraft, ships, AUVs and buoys. Called the Autonomous Oceanographic Sampling Network, the study is seen as a pilot project for future integrated studies of the physical properties of large expanses of ocean over extended periods.

Looking to the future, Glenn and Bellingham want to give teams of AUVs working in concert even more autonomy. With the help of Naomi Leonard of the Dynamical Control Laboratory at Princeton University in New Jersey, they hope to get gliders to act like swarming insects. Using software algorithms that

Slocum gliders can cover huge distances cheaply.

Super subs: Dorado explorers could find use in a range of scientific and military applications.

allow a glider to choose between several different simple behaviours, they envisage fleets of them behaving as if they possess a collective intelligence — surfacing, radioing their position to a central control centre, and then plotting their next moves after learning of the position and behaviour of their fellows. "We will have smart societies of gliders," says Glenn.

Oiling the wheels

Glenn's confidence stems in part from the growing interest of the military and the oil industry in AUVs — which is likely to spur investment and technological progress. Burgeoning oil prices and depleted shallow-sea reserves are seeing oil companies prospecting in ever-deeper waters, where AUVs have distinct advantages. The US Navy, meanwhile, has funded the development of the technology because it is interested in AUVs as mine hunters, and to survey the sea bed near coastlines before amphibious assaults. It is no coincidence, for instance, that Bellingham's Dorado AUVs fit snugly within a submarine torpedo tube — and he has founded a company called Bluefin to market a version of the craft, with the military as his biggest client.

There's certainly room for technological improvement. Extending battery life is a major issue, but cutting-edge technologies cannot always be adapted for use in AUVs. For example, Autosub currently runs on 5,000 standard 'D' batteries, more commonly used to power flashlights or transistor radios. Although advanced lithium batteries are lighter and far more powerful, they are too costly to deploy in risky missions under Antarctic ice.

As well as developing such basics, AUV enthusiasts are working on specific technological improvements that will increase the submersibles' value to academic scientists. Until now, AUVs have generally had to be extensively rebuilt to equip them with the suites of sensors that are required for specific scientific missions. But Dorado and REMUS vehicles now have a payload bay into which instruments can be placed. This flexibility, Bellingham suggests, will encourage oceanographers to dream up applications that AUV engineers haven't thought of. "Then they will really become owned by the scientist," he says.

Griffiths envisions future AUVs spending their entire lives in the ocean, stopping only to charge their batteries and download data. New

versions of the gliders, meanwhile, can already generate electrical energy as they pass through a thermocline, a sudden change in temperature beneath the surface waters in the open ocean — which could allow them to remain at sea for months on end. And at WHOI and MBARI, engineers have developed underwater docking stations for their AUVs.

In the future, such stations could allow AUVs to operate independently from surface research vessels — perhaps sitting moored at key positions in the ocean to be activated when a scientifically interesting phenomenon kicks off. An AUV such as ABE, for instance, could be docked near to an active undersea volcano in the hope of catching it erupting. Only one such event has ever been observed⁴ — by the time a research vessel gets to the site, the action is usually over.

While prototype docking technology is now available, it will require trial and error to perfect. Early efforts to charge batteries at docking stations, for example, caused hydrogen gas to build up in the AUVs' pressurized hulls. "We were turning them into bombs," says Bellingham.

But if the technology develops as anticipated, AUVs could become a powerful extension of another concept on oceanographers' drawing-boards: permanent moored or ocean-bottom observatories connected to the shore by power and communication cables. Small examples already exist, off the New Jersey coast and in Monterey Bay, and the technology could in future allow oceanographers to communicate directly with AUVs over the Internet. Then it will be possible to explore the deep without ever leaving the lab. ■

Tom Clarke works in Nature's news syndication team.

1. Fernandes, P. G. *et al. Nature* **404**, 35–36 (2000).
2. Brierley, A. S. *et al. Science* **295**, 1890–1892 (2002).
3. Rignot, E. & Thomas, R. H. *Science* **297**, 1502–1506 (2002).
4. Fox, C. G., Chadwick, W. W. & Embley, R. W. *Nature* **412**, 727–729 (2001).

Autonomous Benthic Explorer

♦ www.whoi.edu/home/marine/abe_main.html

Autosub

♦ www.soc.soton.ac.uk/PR/Autosub.html

REMUS

♦ www.whoi.edu/home/marine/remus_main.html

Slocum gliders

♦ www.webbresearch.com/slocum.htm

Dorado/Bluefin

♦ www.bluefinrobotics.com

MBARI

Poles apart, or together with Europe?

With Poland about to join the European Union, some of its scientists are eager to be part of the international mainstream. But others still cling to Eastern-bloc traditions. Quirin Schiermeier reports.

C. SOKOLOWSKI/AP

Matthias Bochtler beams with delight. Proudly, the young biophysicist surveys the basement of Warsaw's International Institute of Molecular and Cell Biology (IIMCB). Here, behind a large window, is his group's pride and joy: a brand new X-ray protein crystallography facility, worth some half-a-million euros (US\$500,000). Glancing at a computer monitor, Bochtler nods contentedly — the first data look promising.

In Warsaw, Bochtler is a something of a scientific celebrity. A former PhD student of the German biochemist and Nobel laureate Robert Huber, Bochtler is one of the few foreign researchers pursuing a career in Poland. And he is emblematic of the changes that leading Polish researchers say will be necessary if the country is to reap the scientific benefits of its deepening integration with the West.

Poland is the largest of ten countries that will join the European Union (EU) in 2004, and its scientists are already eligible to participate in EU-funded research. Making the most of these opportunities will require the country's scientists to adapt to the harshly competitive nature of international science. But in a scientific community dominated by a generation trained during the isolation of communist rule and martial law, that won't be easy.

The IIMCB, established in 1997, is a focused effort to put the élite of Polish biology on an international stage. "From the beginning, the idea here was to create something really new; an institute modelled along the lines of the US National Institutes of Health, with a scientific environment open to international competition," says neurobiologist Jacek Kuźnicki, the IIMCB's energetic director.

Attracting high-flying scientists such as Bochtler was central to the plan. Independent, talented lab leaders — from home and abroad — will be crucial to the development of Polish science, says Kuźnicki. With this in mind, the IIMCB's group heads are selected in open competition by an international advisory board. Similarly impor-



Breaking down the barriers: young Poles may back their country's impending membership of the European Union, but reaping the scientific rewards might hinge on Poland attracting more talented researchers like Matthias Bochtler (left) from abroad.



tant, argues Kuźnicki, is the institute's system of fixed-term contracts and stringent evaluations — a radical departure from the job-for-life atmosphere that still pervades most of the 80 or so institutes run by the Polish Academy of Sciences.

"Our continued existence is fully dependent on our scientific success," says Maciej Żylicz, head of the IIMCB's molecular biology department, whose publications on the role of heat-shock proteins in tumour cells have made him one of Poland's mostly highly cited biologists. "I accept that, as a scientist, I am asked to constantly prove that I am good," he says. "This is, however, an unusual thought in Poland, where you are tenured almost as soon as you have received your PhD."

The IIMCB's formula seems to be working. Not only is the institute seen as a model by Poland's State Committee for Scientific Research (KBN), competing successfully to win a generous portion of the country's meagre science budget, it is also attracting grant funding from foreign sources. And for Bochtler, whose work is backed by Germany's Max Planck Society, the IIMCB represents an

appealing scientific home. "I am working on an island of plenty here," he says. "Sometimes, I really think it is unfair that most Polish colleagues must get by with much less. But there is so much friendliness here, so much talent and hard work, and no envy at all."

Cautious approach

Away from the IIMCB, however, things look much less rosy — and some of the initial reactions to its establishment did have the green tinges of jealousy. "There was huge opposition among Polish professors," recalls Kuźnicki. "The Academy of Sciences in particular was extremely sceptical about the principles of external quality control and rolling tenure." Kuźnicki attributes this opposition to a fear among the academy's hierarchy of losing power and influence. But some researchers argue that there are good reasons to be cautious about applying foreign recipes to Polish science.

Włodzisław Zagórski-Ostoja, director of the Academy of Sciences' Institute of Biochemistry and Biophysics in Warsaw, agrees that the academy needs new blood — and at his own institute he has created a Polish-French Center for Plant Biotechnology to

A. SŁIWOWSKI

support the exchange of young scientists. But he believes that Poland is not yet ready for a system based on fixed-term contracts, requiring many scientists to further their careers by moving from lab to lab. "Unrestricted mobility is not the only prerequisite for good science," Zagórski-Ostoja argues. "It is a feature of our society not to hop around so easily."

But other scientists say that Polish science must adapt if it is to compete internationally. "The main problem is that Poland supports institutions, not researchers, and that 90% of Polish professors retire at the same institute where they started their career," says Leszek Kaczmarek, a neurobiologist at the Nencki Institute of Experimental Biology, which shares a campus with the IIMCB and Zagórski-Ostoja's institute.

In the 1980s, Kaczmarek worked in the lab of Renato Baserga at Thomas Jefferson University in Philadelphia, where he studied the molecular basis of learning and memory. Kaczmarek says he is proud that every student who received a PhD in his lab has taken up a postdoctoral position abroad. Indeed, it is estimated that almost half of Poland's young scientists are currently working in the United States.

Michał Kleiber, Poland's science minister and head of the KBN, is relaxed about this statistic — his priorities are boosting science funding and reforming Poland's science base so that the best of the émigrés eventually return. "Numerically, our research base is strong enough, but we are still in a period of transition," says Kleiber.

The supply of young talent clearly isn't a major problem, as student numbers have more than quadrupled since 1990. But the KBN's entire annual budget has shrunk by some 20% since 2000, and now totals just E670 million — roughly equivalent to the annual research spending at a couple of Western research universities. A similar amount

Michał Kleiber: science needs to be a Polish national priority.

East-West divide: science and technology benchmarks across Europe

	R&D spending (% of GDP, 1999)	Annual scientific publications per million population (1999)	High-tech exports (% of total exports, 2000)
European Union	1.93*	755	19.7
Poland	0.75	221	2.1†
Czech Republic	1.24	352	7.8
Hungary	0.69	370	22.9
Romania	0.40	70	4.5
Slovak Republic	0.66	293	4.1†
Slovenia	1.51	577	3.7†

Source: European Commission; *2000; †1999.

comes to Polish labs from other sources, including the EU. As a proportion of national wealth, Poland spends much less than the EU average, and even fails to match Eastern European neighbours such as Slovenia and the Czech Republic (see Table, above).

Kleiber, a physicist and engineer, worked in the United States, Germany and Japan before becoming director of the Academy of Science's Institute of Fundamental Technological Research in Warsaw in 1995. He made what he hopes will be a temporary switch to politics last year, realizing that the future of Polish science depends on it being made a higher national priority. "It is a tremendous task to convince people here to invest in science rather than worry about tomorrow," Kleiber says.

Uphill struggle

But convincing his political peers that investment in research will yield economic returns will be tough — the country's traditional strengths in applied physics, chemistry, mathematics and engineering all lack a domestic industry to absorb their results. Poland's weak industrial base is similarly limiting its opportunities in EU-funded research, Kleiber admits.

But he argues that his fellow ministers are starting to listen. "My government colleagues accept the importance of science, not least with regard to EU membership," says Kleiber. One positive sign, for instance, is that the KBN will later this year be transformed from a state committee into a fully fledged ministry, giving it more clout within the corridors of power — a change that Kleiber hopes will be converted into more coherent science policies and improved funding.

So far, however, governmental goodwill has not translated into hard cash. "If you subtract obligations to miners, shipbuilders, farmers and so on, there is little left for science," says Kleiber. Given this background, the KBN is looking for ways to concentrate available funds on the most competitive research units. It has compiled a list of 157 Polish labs that have applied for EU funding and were rated as 'good' or better by Brussels' peer reviewers.

The plan is to priori-

tize these labs for future funding. But releasing sufficient cash to allow top Polish groups to compete on equal terms with Western competitors would mean withdrawing government backing from poorly performing labs or institutes. And in this regard, progress is sluggish. Jan Krzysztof Frackowiak, under-secretary of state at the KBN, estimates that only a quarter of the total volume of research carried out at institutes of the Academy of Sciences is internationally competitive. But currently only one academy institute, for ecology, is earmarked for closure.

It's a similar story at the universities. At the 700-year-old Jagiellonian University in Krakow, the academic cradle of Nicolaus Copernicus, there are positive signs. Last October, it opened Poland's first university biotechnology centre, which is already collaborating with prominent Western groups and in the autumn will launch English-language courses. But such initiatives are the exception in a system that remains insular and uncompetitive.

Current employment laws and powerful trade unions restrict the KBN's ability to close academy institutes. And reducing the number of staff in universities isn't an option: they are needed to teach the growing numbers of students.

Kleiber hopes that once Poland joins the EU, its scientific infrastructure will be boosted by money from the EU's Structural Funds, designed to develop the economy of the union's poorer regions. But in the absence of significant progress in transforming the country's industrial base, winning a sizeable proportion of the separate funds available under the EU's research budget will require Poland to raise more of its research centres to the internationally competitive level of the IIMCB. And it is difficult to see that happening without widespread closures of labs that are never likely to make the grade.

For Polish science, the road to successful European integration could be long and rocky.

Quirin Schiermeier is *Nature's* German correspondent.

International Institute of Molecular and Cell Biology
♦ www.iimcb.gov.pl
State Committee for Scientific Research
♦ www.kbn.gov.pl/en
Polish Academy of Sciences
♦ www.pan.pl/english

Venezuelan researchers call for international help

Political conflict is endangering human life, scientific endeavour and a rich environment.

Sir — Venezuela has entered 2003 deadlocked in a dangerous political confrontation and a general strike that has lasted two months so far. Scientific endeavour, like everything else, has been seriously disrupted by the crisis, partly because many researchers and post-graduate students support the strike, but more imminently because salaries, project funding and basic services such as gas, electricity, liquid nitrogen, transport and the Internet are no longer reliable. We find ourselves in a chaotic environment where it is next to impossible to fulfil national and international research commitments.

One notorious casualty is Intevep — the R&D branch of the national oil company PDVSA — which has practically come to a standstill and so is now earmarked for 'major restructuring' by the Ministry of Energy and Mining.

Our Venezuelan Institute for Scientific

Research (IVIC) and the national research universities have been prostrated by financial cuts starting last year. We hear many threats of government intervention.

The Association of Investigators of IVIC (AsoInIVIC) has been hosting assemblies and electronic debates in the past two months. We have sent out press releases in which we deplore the recent spread of politically motivated violence, including the killing of unarmed demonstrators by paramilitary groups.

A consultative referendum was legally scheduled for early February, but has been suspended by the Supreme Tribunal. A table of negotiation and agreements has been set up by the Organization of American States, chaired by its Secretary General, Dr César Gaviria: we have urged the 20 negotiators to support the consultative referendum.

We have also issued warnings against

misuse of PDVSA's equipment and infrastructure by uncertified personnel because accidents involving human lives and the environment have already occurred — for example, 79 reported oil spills on Lake Maracaibo since the conflict began.

We at AsoInIVIC are struggling to preserve Venezuela's science and its natural environment, which is rich in biodiversity. We ask the scientific community to help by exerting pressure, for example through learned societies and UNESCO; by raising funds to help us meet commitments abroad; by backing local groups, through environment agencies, in denouncing the ecological disasters; and by remaining patient while we return to normality.

C. Mendoza (president), J. A. Urbina (secretary general)

Association of Investigators of IVIC,
Venezuelan Institute for Scientific Research,
PO Box 21827, Caracas 1020A, Venezuela

No reduction in risk of a massive asteroid impact

Sir — An important, technically valid analysis of meteoric fireball data (P. Brown *et al.*, *Nature* **420**, 294; 2002) has been transformed into media stories asserting (falsely) that the threat from dangerous asteroids is less than once thought. *Nature* joined the media hyperbole with words on its cover about "reassessing the hazard", although the data published by Brown *et al.* have nothing to do with the hazard.

Brown *et al.* studied satellite and infrasound data over an 8.5-year period pertaining to impacts of meteoroids 1–10 metres in diameter. Except for the few per cent that are strong nickel-iron, such objects burn up harmlessly high in the Earth's atmosphere. Thus they constitute no direct danger. Only objects perhaps 1,000 times more massive than the largest studied by Brown *et al.* actually reach the ground with much cosmic velocity intact.

Yet many of the media reports about the paper asserted that this implies that the frequency of dangerous near-Earth asteroid (NEA) impacts is decreased, and therefore that the hazard has also declined significantly. In just one of many examples, the *Los Angeles Times* editorialized on 30 November 2002 that "if *Nature* is to be believed" then asteroids capable of causing catastrophic damage "can recede from the worry list".

But the danger from NEAs is not less than once thought. *Nature*, in its own press release, mentioned the extrapolation

by Brown *et al.* of their data to objects tens of metres in size, like the small NEA that caused the roughly 10 megaton 'Tunguska' explosion over Siberia in 1908. But the lower, once-a-millennium estimate for Tunguska derives from telescopic Spaceguard Survey data reported during the past two years (see, for example, D. Rabinowitz *et al.*, *Nature* **403**, 165–166; 2000), not from the study by Brown *et al.* of much smaller objects. And even Tunguska contribute very little to the hazard, which is dominated by rare impacts of NEAs several thousand times more massive than Tunguska (with diameters 1–3 km).

The estimated frequency of such larger NEAs (and hence of the total impact hazard) is not changed by the work that Brown *et al.* have done. Indeed, data analyses during the past year suggest that NEAs of civilization-threatening sizes may be more numerous than previous estimates by tens of per cent.

For such low-probability events as NEA impacts, the important issue is not the statistical frequency but the determination of whether there is (or is not) an asteroid that will collide with Earth in the next few decades. That is the purpose of the Spaceguard Survey — to find these NEAs one at a time, and determine from their orbits whether or not they pose a danger in the near future.

Clark R. Chapman*, David Morrison†

*Southwest Research Institute, 1050 Walnut Street,
Boulder, Colorado 80302, USA

†NASA Ames Research Center, Moffet Field,
California 94035, USA

Was visionary Seaborg from far-sighted stock?

Sir — In a letter that I received inviting me to become a 'Consulting Fellow' for the World Innovation Foundation, I discovered (amid considerable other hyperbole) that some honours are so great as to defy accurate description.

Glenn Seaborg was founding president of the World Innovation Foundation, so it was only fitting that the letter should try to capture the significance of his achievements and his stature as a person: "As one of the twentieth century's greatest scientists and the only person ever in the history of the world to be named after one of the Universe's Elements (Element 106 Seaborgium) whilst still living, [Nobel Laureate Glenn Theodore Seaborg] has left a huge scientific legacy for us all."

Glenn's parents must have been unusually prescient. Fortunately, the website (www.ineed.easynet.co.uk/wif/mainmembers.html) has it right, so only those of us honoured with an invitation letter might have momentarily mused over the order of events.

A. Richard Palmer

Department of Biological Sciences,
University of Alberta, Edmonton,
Alberta T6G 2E9, Canada

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited for length.

Is a picture worth 1,000 words?

Exciting new illustration technologies should be used with care.

Julio M. Ottino

A quick look at *Nature* or *Science* immediately shows how the role of images in science has increased over the past 20 years. It is now rare to find a research article that has no images. Indeed, it is hard to remember that colour is a relatively recent addition to scientific publishing. This increase in illustrations is undoubtedly due to the widespread use of computer graphics — images can readily be enhanced, modified and morphed. It is now relatively easy to draw figures with shadows and multiple reflections between mirror-like surfaces.

Although scientists may like to think they are immune to figures, this resistance may be a remnant of past thinking that has portrayed visual-aided arguments as less than rigorous. Physics and mathematics have been held as primary examples of domains where images play no role. But this view is far too narrow — visual imagination is a central element of scientific imagination¹. Seeing and representing are inextricably linked to understanding. Galileo's sketches of the Moon (Fig. 1) are possibly the most celebrated example. But there is little doubt that the role of images has changed significantly since Galileo's times. The question is, are we now getting ahead of ourselves?

Figures fall into several categories. At one extreme there are those that convey data; at the other, scientific illustration. When

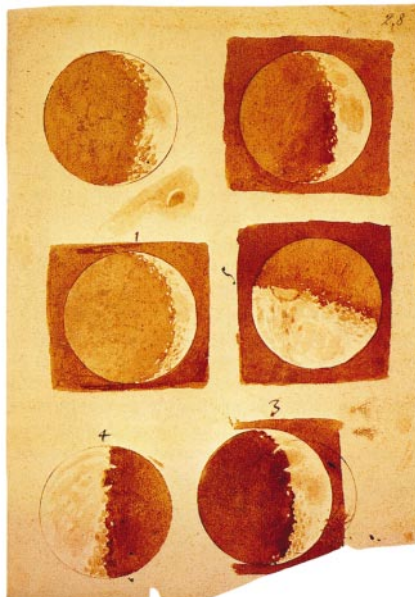


Figure 1 Pictures in perspective: Galileo's sketches of the Moon are more than mere illustrations — they convey relevant scientific information.

images are 'the' experimental result, photo editing opens up the distinct possibility of overmanipulation and misrepresentation. The hard line, that any unreported manipulation constitutes fraud, is short-sighted. If shades of grey in an image range in levels from 134–147, they would all look the same

to our eyes. But rescaling these pixels so that 134→0 and 147→255 readily reveals the differences to the human eye. Well-established tools² such as deconvolution (used, for example, to interpret fluorescence images) enhance the human ability to 'see'.

My objective is, rather, to tackle the seemingly less scientific topic of scientific illustration. Such images enter the equation, for example, to depict mechanisms: how a cell is attacked by a virus; when a DNA helix is unravelling; or when an asteroid plunges into Earth (Fig. 2). Representations of how something might or could be built fall into this area — the many sketches of Leonardo, or more recent examples such as the 'nanolouse' and nanocircuits of Figs 3 and 5. At a more basic level, scientific illustration shows how something is — the Moon in Galileo's times — or how it can be imagined — Bohr's atoms, or how two molecules might interact.

Figures used to be part of the thought and discovery process. For Leonardo, Galileo and mathematicians such as Riemann, the image was part of the thought process^{3,4}. The same instrument penned words and drew lines. This seamless integration is more than a quaint sentimental point. A line is tentative; in a line drawing one hears the voice of the author: "this is what I think happens..." or "this is how I imagine this mechanism works". One can see a mind at work, switching from word to image (Fig. 4). It is hard to argue that replacing Leonardo's drawings in the codices with computer-gen-

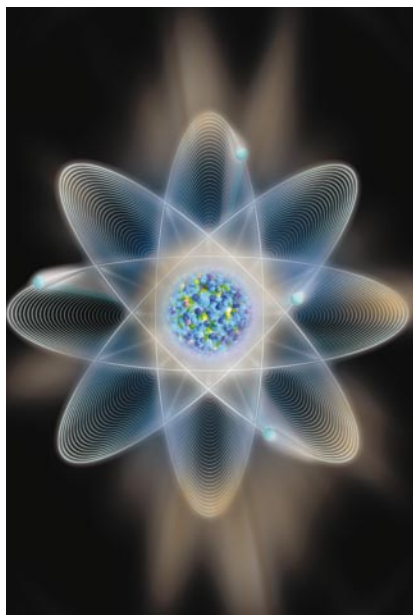


Figure 2 More real than reality: artistic impressions, from atoms to asteroids, can be striking, but do they aid scientific understanding?



Figure 3 Could this device function? This award-winning computer-generated image of a nanolouse initially looks realistic, but close inspection reveals an element of 'artistic licence'.

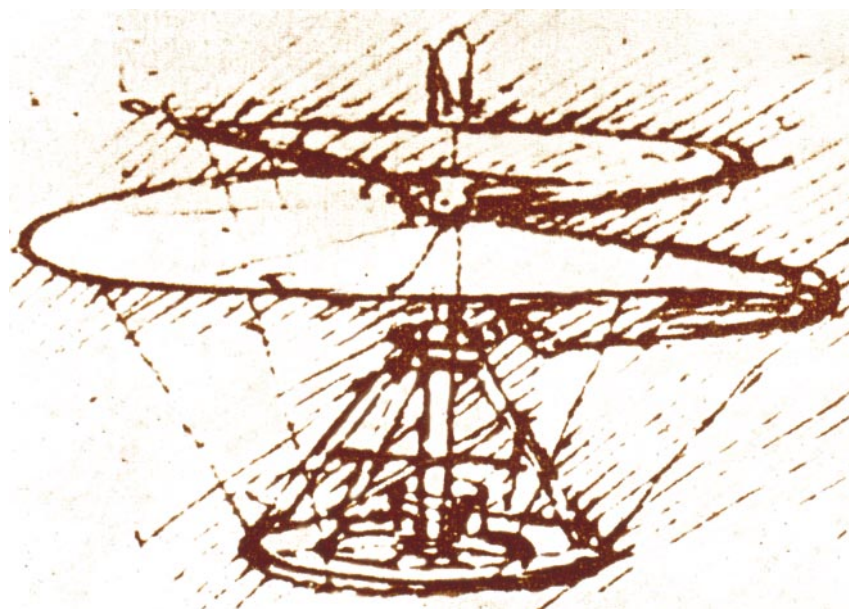


Figure 4 Helicopters and HIV: Pictures can be part of discovery and thought processes, providing an insight into how devices may operate.

erated images is really progress.

These days, of course, text and figures are handled in different ways. In many instances, figures are left in the hands of artists and illustrators (Fig. 2). There is a real concern that the practice of using artists' impressions or mock-ups denies the physics of the situation, or is so convincing that the line between fantasy and reality is blurred. Examples abound in science, most lately in the emerging discipline of nanotechnology.

Rules of the game

This all means that publications should establish guidelines of what manipulation or enhancement is permissible. I believe that we need to evolve a system that is the scientific equivalent of that for films, analogous to David Lean's expansion of "directed by" to "edited and directed by". Attribution and clarification of image manipulation⁵ will make the figure truly 'scientific'.

A sensible first rule would be that pictures must not be divorced from science and scientific plausibility. Images should not conflict with or violate known physics. Of course, one should not take this too far — it is pointless to criticize the circular orbits in Bohr's model of the hydrogen atom or the practice of colouring atoms in molecular models.

It is hard to pick good examples of bad pictures in this regard. Poor images do not tend to receive wide exposure. We are looking for pictures that have been singled out as being particularly good, even dazzling, artistic depictions, perhaps selected for covers and art prizes — images expressly picked on artistic merit, but also for attracting attention. In illustrations, the observer generally assumes that the level of information in the figure is uniform — if something in the figure is intended to be 'realistic', then the rest

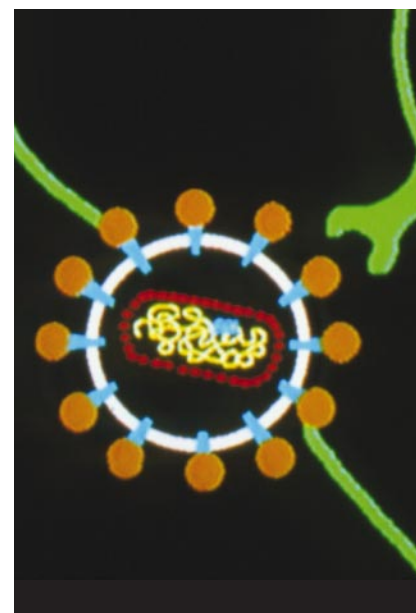




Figure 5 A nanocircuit from the cover of *Science*. This picture purports to offer scientific insight — but if the carbon atoms are visible, then the much larger gold atoms in the structure should also be on view.

of the figure should be as well. If this is not the case, the caption should make this clear. For example, in the cover figure of the nanocircuit in Fig. 5, as one can see carbon atoms in a nanotube, one should also see gold atoms in the support (ratio of masses more than 10, ratio of diameters about 2). An experimental figure in the paper⁶, which the cover illustration purports to represent, shows gold atoms whereas the nanotube itself can barely be discerned. The necessary requirement that no physics is violated is broken by Fig. 5, which shows shadows, reflections and so on as they would occur in everyday life, which most definitely do not occur in quite the same way at the nanoscale.

Aesthetic rules would also be useful. Images, just like text, should not be more complicated than they need to be. The image in Fig. 6, for example, is clearly not intended to be confused with reality, it simply purports to show the coupling between two molecules that can give rise to new spectral features. However, is it really necessary to have so much detail: full, three-dimensional solid-like arrows with shadows, when just a few lines will do or when the molecules themselves are depicted in a clearly schematic way?

Another rule is needed to address the extrapolation of everyday physics to molecular and mesoscopic scales. We should resist the temptation to prejudge answers, recalling Richard Feynman's 1959 challenge to build a very small motor, and its realization only a year later. This rule needs to be particularly clear and explicit in dealing with images of what must be imagined and cannot be seen. Pictures often precede devices; they indicate how something might look without it having been built, which can



Figure 6 *Science* as art. Clearly not intended to be a realistic portrayal, this image of two molecules is unbalanced as it includes too much unnecessary detail in arrows and too little on the molecules themselves.

influence how the effort might actually be attempted.

Consider the beautiful image of the 'nanolouse' — the micromachine with significant functionality which is the size of a red blood cell (Fig. 3). This image won first prize in the 'Science Concepts' section of the 2002 Visions of Science Awards. The image looks so real that it is easy to imagine a viewer being fooled into believing it has already been built. This is important in terms of public reaction, especially in the backdrop of scenarios such as in Michael Crichton's new novel *Prey*, which portrays swarms of self-replicating nanomachines destroying all other life forms that they encounter.

But could the nanolouse in Fig. 3, in fact, be real? A bit of 'physics' thinking helps. The image itself shows features such as tubing systems that are about 200 nm in size. One would never see this level of detail in a light microscope (features at this scale get fuzzy, as 200 nm is the Rayleigh resolution limit). In addition, colour would not show up in a scanning electron microscope. So, if this picture is 'real', it must also be heavily enhanced and manipulated.

The important question concerns the device itself. The nanolouse manoeuvres to get close to a red cell, then grabs it with glass-like pointed claws, and injects a needle.

Pictures must not be divorced from science and scientific plausibility.

The picture projects an aura of planning, purpose and serenity, very much like the space shuttle docking with a space station or a crane grabbing a steel beam. But life is not serene or deterministic at these length scales. The size of red blood cell is 7 μm . Below these scales, thermal motions provide a randomizing effect and the tips of the pincers may fluctuate. Surface forces, such as van der Waals forces, electrostatic forces and perhaps the interactions between the device and the cell would be critically important. If both the cell and the glass were negatively charged, for example, the cell would repel the claws. A very sophisticated control process would be required to position the claws around the cell, especially when considering random brownian contributions and attractive or repulsive forces. How would the device get this information? Processing information and injection both require energy — where would this come from? How is the micromachine propelled? As the effects of viscosity completely dominate dynamics at these scales⁷, the mechanisms of propulsion are unlike those that work at macroscales⁸.

Figures influence people, sometimes subconsciously. Would an image such as Fig. 3 influence someone trying to design a nanolouse? If the objective is to design a micromachine to inject a substance into a red cell, the resultant device would not look like the one portrayed in the figure.

Finally, scientists publishing figures as part of their research papers should always ask some general questions. What is the point of the image? Is the objective to teach, to excite or to show how things could be? How can this objective, whatever it might be, be made clear to the viewer? There are many new tools for making beautiful drawings, but if good use is to be made of them, scientists and artists should collaborate closely. Going all-out with computer-generated images without asking questions like those discussed here may be a perfect example of confusing progress with progression. ■

Julio M. Ottino is at the R.R. McCormick School of Engineering and Applied Science, Northwestern University, Evanston, Illinois 60208, USA.

- Holton, G. *Daedalus* 125, 183–208 (1996).
- Misell, D. L. in *Practical Methods In Electron Microscopy* (ed. Glauret, A. M.) Vol. 7. *Analysis, Enhancement and Interpretation* (North-Holland, New York, 1978).
- Hadamard, J. *The Psychology of Invention in the Mathematical Field* (Princeton Univ. Press, 1945).
- Poincaré, H. *The Foundations of Science* (Univ. Press of America edition, Lanham, MD, 1982).
- Frankel, F. *Envisioning Science: The Design and Craft of the Science Image* (MIT Press, Cambridge, MA, 2002).
- Bachtold, A., Hadley, P., Nakanishi, T. & Dekker, C. *Science* 294, 1317–1320 (2001).
- Purcell, E. M. *Am. J. Phys.* 45, 3–11 (1977).
- Becker, L. E., Koehler, S.A. & Stone, H. A. *J. Fluid Mech.* (submitted 2002).

Acknowledgements

I would like to acknowledge discussions with Philip Ball, Howard Stone, Mark Ratner, Darell Velegol and Joe Zasadzinski.

Stress factors

The science of our flexible responses to an unpredictable world.

The End Of Stress As We Know It
by Bruce McEwen

Joseph Henry Press/Dana Press: 2002. 256 pp.
\$27.95, £20

Does Stress Damage the Brain?
Understanding Trauma-related Disorders from a Mind-Body Perspective

by J. Douglas Bremner

W. W. Norton: 2002. 311 pp. \$30

Thomas Elbert & Brigitte Rockstroh

Which organ do you value most? It ought to be your brain, especially as you enjoy sitting here reading *Nature*. Are there ways of protecting or perhaps even improving the functioning of this precious organ? The two books reviewed here could potentially change the way you think about your brain and the way it functions. *Does Stress Damage the Brain?* by psychiatrist and neuroscientist J. Douglas Bremner is an account of brain functioning and stress that adds facts, depth and detail to the subject. A more conceptual and broader scope is taken by the other book, *The End of Stress As We Know It* by Bruce S. McEwen, a pioneer and leading scholar in research on stress and 'allostatic load' — a term he uses to describe the cumulative wear and tear of life. Stress, allostasis and their relationship with disease or mental well-being within the frame of life are the topics of McEwen's book.

The body, including the brain, can deal with danger in a flexible and adaptive way. In contrast to homeostasis — an organism's ability to maintain a steady internal state — allostasis refers to the flexibility in adjusting to stressors that range from Hans Selye's types of physical deprivation (such as cold, noise, deprivation of food and sleep) to fear-provoking situations that trigger an alarm response, even through imagination.

The Greek word *allo*, meaning variable, is used to emphasize this ability to choose from an arsenal of attack and defence armaments to counter a negative impact. Even with just minor cues, the brain can activate any appropriate alternative of the 'flight–fight–freeze' defence cascade. Fight, according to phylogenetic origins, refers to the transfer of maximum energy to the parts of the body that need it most. However, a defence system that has its roots in an archaic fish can be absurd in a modern human. Does it really serve your best interests to supply additional blood and oxygen to your leg muscles when you're driving a sports car? Is it necessary for stored carbohydrates to be liquidated into blood sugar when discussing issues with the ethics committee? No, of course not.



Fight or flight: stressful situations trigger the alarm response, here shifting energy to the leg muscles.

Allostasis has evolved as the response for running away from a predator, escaping acute danger or fighting off a threat. It also provides the modern human with higher levels of stress hormones in the morning so we can meet the challenges of the day. If the allostatic load becomes too high — a state defined as the permanent initiation of warding off stress (or being 'stressed out') — disease will knock at the door in the form of aches and pains, loss of appetite or over-eating. A long-term high allostatic load can also damage organs, including the brain.

It is the treatment of allostatic load that makes McEwen's book such a fascinating read. It is not so much the wealth of facts that have been brilliantly summarized in an often entertaining manner, but the concise and sharp conceptualization of the findings that show the reader how to improve both physical and mental health. McEwen artfully arranges the many pieces of the physiology of stress into a holistic mosaic with a paradigmatic leap towards holistic medicine. It is this leap that makes this work useful not only to the reader who is unfamiliar with the topic, but also to a reader who is aware of all the individual facts, but who perhaps has never bothered to compose them into the picture of allostasis.

Allostasis, explains McEwen, begins deep in the brain, where the hypothalamus uses two routes to sound the alert to the adrenal glands, far away at an outpost above the kidneys. The adrenals respond by pouring out major stress hormones. The hypothalamic–pituitary–adrenal (HPA) axis is the elite force of the defence cascade; when functioning

properly it helps us to deal with crisis. When unremitting stress forces the axis to tilt one way, the result can be anything from a long-lasting head cold to depression. When tilted the other way, towards a flattening of the rhythm of stress hormones, undesirable consequences may be abdominal fat, loss of muscle mass and mental ailing. When the danger is over, the stress response shuts down — at least in wild animals. Humans, however, seem to be unique in that they can keep the HPA axis going indefinitely. The hormones ultimately make their way back to the brain, affecting both behaviour and health.

Bremner's account addresses atrophy of the hippocampus as a result of stress, a phenomenon that can make the individual more vulnerable to life-threatening encounters. Two prime targets for stress hormones in the brain are the hippocampus and the amygdala. In these structures, neurons may change shape, and connectivity and even the number of cells may be dramatically altered.

The journalists' questions of 'when, where, what and who?' are the territory of the hippocampus. The emotional impact is added by the amygdala and, through its connections, the frontal cortex. When stress levels are very high, the hippocampus may be damaged, and its essential role in forming memories, particularly those for context, is at risk. In dangerous situations, it is important to remember one context, not many. Because of this, the hippocampal formation becomes more specific through pruning, while at the same time strengthening the amygdala's fear networks.

Traumatic stress may cause even greater

brain damage, so symptoms that are related to psychological trauma can be understood as neurological disorders. In his book, Bremner reviews the neuronal, hormonal, autonomic, mental and behavioural bases of allostatic load brought about by traumatic experiences.

Like McEwen, Bremner rehearses well-known facts, but he also conveys an insight that slowly arises to a revolutionary thinking in medicine and psychiatry: that stressful experiences differentially activate a variety of responses designed by evolution to counter danger. The different chemical messengers may cause deficits in hippocampus-based learning and memory, and their effects on the amygdala and medial prefrontal cortex may lead to impaired inhibition of fear responses. As a result, emotional and autobiographical memory may become fragmented. In addition, repeated or chronic exposure to traumatic stress may lead to long-term dysregulation of these systems, resulting in impaired functioning and symptoms of stress-related disorders such as hyperarousal, dissociation, flashbacks, avoidance and depression.

The benefit of both books (perhaps more for clinicians and educated laypersons than for neuroscientists) is in the integration of well-known biological evidence from a 'mind-body' perspective. The strength of McEwen's book is its broad conceptualization, whereas Bremner's book emphasizes experiments from his own laboratory. Both volumes include many personal and clinical observations, making them not only educational but also enjoyable to read. This is a good thing; after all, you wouldn't want to increase your allostatic load — that could be bad for your brain. ■

Thomas Elbert and Brigitte Rockstroh are in the Department of Psychology, University of Konstanz, D-78457 Konstanz, Germany.

Sowing the seeds of modified crops

Genes for Africa: Genetically Modified Crops in the Developing World

by Jennifer A. Thomson

University of Cape Town Press: 2002. 192 pp. R120

Gordon Conway

This is a gem of a book. It is clear and concise, it makes the complex seem simple without losing the essential truths, and, as far as I can tell, it is accurate, with no innuendo, no half-truth and no wild extrapolation.

Jennifer Thomson is professor of microbiology at the University of Cape Town in South Africa and has been working in biotechnology for more than 20 years. She is



Genetically modified food could be just one element of an integrated solution to Africa's food crisis.

unashamedly in favour of genetically modified (GM) crops and takes delight in demolishing the wilder scaremongering of the anti-GM activists. As she points out, we need to maintain the distinction between a hazard and the probability of a hazard occurring. It is easy to dream up potential hazards — the question is, how likely are they? And, as she effectively argues, in most instances the chance is very remote indeed.

In just 190 pages Thomson covers a wide range of topics: the biological basis of genetic engineering, the current status of GM crops, food safety, biodiversity, patenting, labelling and so on. Two appendices cover the testing of GM foods for allergies and the likelihood of horizontal gene transfer. She defines every technical term as it is introduced, and uses diagrams, boxes and colour illustrations to make her points.

I agree with Thomson that all the evidence suggests that GM crops are as healthy for humans as are conventional or organic crops — and, in some respects, are even healthier. The remaining concerns centre on the probability of gene flow, providing wild relatives with competitive advantages that could significantly change natural ecosystems. I think the jury is still out on this, and I feel that she could have expanded further on the topic (but then I am an ecologist).

The book should be read by anyone interested in crop biotechnology, but its special significance is in relation to Africa and that continent's desperate effort to achieve food security. In Africa there would not be enough food to go round even if it were evenly distributed. Average cereal yields are a pitiful one tonne per hectare, compared with nearly three in the developing countries of Asia. Yields have to rise to two tonnes at least — in bad years as well as good — for African farmers, who typically farm a hectare or less,

to be able to feed their families and to grow crops for cash with which to educate their children and buy medicines.

As Thomson readily acknowledges, GM crops do not constitute a magic bullet. Food security will only come from a combination of better input and output markets, access to cheap fertilizers, better water and soil management, and improved crops. Crops can be improved by conventional breeding, by tissue culture and by marker-aided selection. But for many improvements — such as drought and salinity tolerance, enhanced phosphorus and nitrogen uptake, resistance to the parasitic weed *Striga*, resistance to viruses and bacteria, delayed ripening of fruit and vegetables, and improved amino-acid content of forage crops — GM technology is going to be essential.

The terrible famine now affecting southern Africa and Ethiopia makes this a timely book. This is not to argue that GM technology will make famine a phenomenon of the past. But it is Thomson's contention that in the long run, by using genes from indigenous African plants, such as the resurrection plant *Xerophyta viscosa*, we may eventually end up with crops that can withstand Africa's devastating droughts.

During these past few months, pressure from northern anti-GM activists has compelled several African countries to reject food aid for fear that the grain might contain some GM seed that might then be planted. Zambia even rejected food aid when the maize was offered as milled grain. The European Union pronounced the grain safe and the US government could not see why it should not provide grain that is eaten by Americans virtually every day of the week.

As Thomson argues, Africans must be free to choose whether to develop and grow

GM crops, without such strong-arm tactics from either side. They need the kind of clear, authoritative information that she provides so that they can make up their own minds in a logical and dispassionate manner.

I recently had a long conversation with President Museveni of Uganda. He asked many thoughtful and penetrating questions about GM technologies. After our talk I sent him a copy of *Genes for Africa*. I know it will have given him many of the answers he is seeking. ■

Gordon Conway is president of the Rockefeller Foundation, 420 Fifth Avenue, New York, New York 10018-2702, USA.

..... A knotty story

Knots: Mathematics with a Twist
by Alexei Sossinsky, transl. Giselle Weiss
Harvard University Press: 2003. 160 pp.
\$24.95, £16.50, £24.95

Andrzej Stasiak

Not surprisingly, knot theory is a complicated business. For instance, the common knots encountered in daily life are not even considered to be knots by mathematicians. A tied shoelace is not knotted, in mathematical terms, because it can be untied by a continuous deformation. But the situation changes completely if the two loose ends of a shoelace or a piece of rope are permanently joined together. Now, even if the newly spliced rope forms a simple circle, a mathematician would immediately recognize it as a knot, albeit a trivial one, as it is no longer possible to untie it with a continuous deformation; this would require cutting the rope. It is these mathematicians' knots — closed curves in three-dimensional space — that are the subject of Alexei Sossinsky's *Knots*, which was first published in French in 1999 and is now available in English.

Sossinsky, a professor of mathematics at the University of Moscow and a knot theorist, gives equal attention to both the historical and the theoretical aspects of his subject. He begins in the 1860s with Lord Kelvin's supposition that atoms are formed of closed vortices of ether, and that what set apart the atoms of different elements was the different ways in which those vortices were knotted. Kelvin's idea won its share of followers and launched scientific interest in the theory of knots, resulting in a systematic classification of various knot types. As it turned out, however, atoms are not knotted (at least not in the simple form imagined by Kelvin), and physicists relegated interest in the subject to mathematicians who might be less concerned with its utility and applications.

In recent years, however, there has been an intriguing convergence, as the mathematics of knots and the intricacies of quantum



It's knot unusual: like other knots recognized by mathematicians, the trefoil knot is a closed loop.

physics have become, well, knotted together. Sossinsky's final chapter is placed in the near future (2004), as this may be the time when a worthy successor to Kelvin demonstrates a deep connection between knots and elementary particles.

Sossinsky concentrates on the central problem in knot theory: how to tell whether two closed curves in space are topologically identical. The first step in the identification process is to assess whether the two knots can be twisted, bent, shrunk or stretched — all continuous deformations — into an identical form. If they can, the two knots are topologically identical. The catch is that even if you cannot make one knot into the same form as the other, it still doesn't mean that they are not the same type of knot — perhaps you have simply failed to come up with the appropriate sequence of topological deformations. To solve the problem, knot

theorists have invented various mathematical tools that can distinguish different knot types irrespective of their actual three-dimensional shapes.

As Sossinsky describes it, the most intriguing tool is based on the discovery that knots that can be converted into one another by a simple set of surgical operations (cutting and pasting) are related to each other in a specific mathematical way, regardless of their actual shape and topology. For the reader, it is at first strange and puzzling to see equations that use projections of these surgically related knots as graphical symbols. Still, with Sossinsky's help and with a little work, you will be able to write and solve these same equations for other sets of knots. This experience alone, if you're willing to put in the effort, makes the book worth reading.

As a biologist, I was pleased to see that Sossinsky compares surgical operations at knot crossings to the action of DNA topoisomerases — enzymes that change the topology of DNA molecules in living cells. Indeed, passing the overlying segment through an underlying one (and vice versa) corresponds precisely to a reaction mediated by DNA topoisomerases. However, Sossinsky is wrong to assert that a topoisomerase can eliminate an individual crossing; this requires another enzyme, a resolvase, which mediates the process of site-specific recombination. Sossinsky could have avoided this and other mistakes by consulting a biologist while writing about DNA. ■

Andrzej Stasiak is in the Laboratory of Ultrastructural Analysis, University of Lausanne, Lausanne-Dorigny CH-1015, Switzerland.



What's blue and sticky?

The toes of the Tokay gecko, shown above, are worth a closer look because they can adhere to the smoothest surfaces. But this isn't the only neat feat that reptiles can perform. It is well known that chameleons can change colour for camouflage, and many lizards can pull off the

handy trick of growing a new tail if their old one is, well, pulled off. For a more comprehensive, and beautifully illustrated, look at reptiles, see *The New Encyclopedia of Reptiles and Amphibians* (Oxford University Press, £25), edited by Tim Halliday and Kraig Adler.

Solid foundations

Summarize yourself in the form of a title of a paper in Nature.

Spectrum of energy release by attrition and enlightenment in the cracking of geological quandaries.

What was your first experiment as a child (on pets/siblings/dead flies, and so on)?

Mixing chlorate sold as weed-killer with sulphur and sugar, and then striking a match. The crackling sound is heaven and the sweet smell unforgettable. I do not recommend conducting the experiment on the wooden floor of an attic amid piles of old books. There is a benevolent god for chemistry apprentices.

Whose graduate student would you most like to have been (historical impossibility notwithstanding)?

Harold Urey, who retired from the Manhattan Project to invent geochemistry and lobby for the Apollo project. Almost every PhD student that he had left an indelible trail of genius.

What single scientific paper or talk changed your career path?

With a couple of slides on mantle tomography, my seismologist friend Rob van der Hilst shattered the 20-year-old paradigm of two-layer mantle convection. "We must have been wrong all along," is not a good sentence to hear reverberating in your head.

What book has been most influential in your scientific career?

Conduction of Heat in Solids by H. S. Carslaw and J. C. Jaeger (1959). I cannot find a better example of a single book that encompasses the huge variety of modelling methods and perspectives as applied to a simple problem. Its presentation is, of course, outdated — but the same can be said for Newton's *Principia*.

What's your favourite conference destination, and why?

The autumn meeting of the American Geophysical Union in San Francisco combines the best talks in Earth sciences, fabulous scenery, a unique atmosphere merging all the cultures from around the Pacific and (please do not tell the French) a cuisine that makes my taste buds shiver with delight.

What was the worst/most memorable comment you ever received from a referee?

For my first paper in *Nature*, the core referee's objection was not the data, not the science, but that my lab had not established its reputation in the field.

What book is currently on your bedside table?

I have just turned the last page of *The Mating Mind* by Geoffrey Miller. OK, I am French. But this brought to an end my personal Dark Age about life, evolution, art, politics and the whole game of men and women.

What music heads the playlist in your car or lab?

Give two young guys in Vienna a keyboard and they will make seven generations dream that they know what heaven is about. What had I done at the age Mozart and Schubert died?

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

I would drink the 20-year-old father of group theory, Evariste Galois, under the table so that he would miss his fatal duel the next morning. His love affair with the young Stephanie prompted her fiancé to shoot Galois dead and stall a particularly important branch of mathematics for decades. Galois had spent the night before the duel writing a major piece on equation theory.

Where and when would you most like to have lived or worked?

Between 1947 and 2003 in the Western world, which is living through its golden age of freedom and affluence.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do? Praising or bashing, I will move to the back of the plane and smile in contentment.

What's the best piece of advice you've ever received?

"Young man, there is no one good reason to keep smoking." (Clair Patterson, Pasadena, 1977.)

What do you most dislike about having research published?

Publication prematurely freezes the thinking process.

The Internet is the bane of scientists' lives because...

...it blurs the limit between grey literature and professionally prepared publication.

What one thing would you rescue from your burning laboratory?

When an explosion took part of the roof off my lab last September, I rushed through the smoke to save the beakers containing the dissolution of martian meteorites. Is there anything of more worth than a frag-



Francis Albarède

Francis Albarède is professor of geochemistry at the Ecole Normale Supérieure in Lyons, France, recipient of the Bowen Award and the Holmes medal, and senior editor of the *Journal of Geophysical Research: Solid Earth*.

ment of another planet collected through the unrelenting efforts of many across the hostile deserts of Antarctica and Sahara? Maybe the pyramids or *Mona Lisa*, but they are not allowed into my lab.

What do you do to relax? (Please bear in mind that this is a family journal.)

Think — the ultimate luxury of modern life.

Harry Potter or Lord of the Rings?

I have read Harry Potter only once whereas I have lost count of how many times I have trudged along with Frodo. With *The Lord of the Rings* I dream with my eyes open. And it is a true literary masterpiece: Tolkien is the Victor Hugo of our reverie.

What would you have become, if not a scientist? A talented and inspired pianist.

Name one extravagance that you can now get away with because of your eminence.

I can finally turn into a law-abiding, reliable and occasionally generous citizen. Which grad student or postdoc can afford such a luxury?

What music would you have played at your funeral?

I always keep my favourite music to myself. Please bury my favourite CDs and a player with me, just in case. Don't forget the batteries. ■

Magic moments

A. P. Ramirez

Physicists who study condensed matter are interested in the properties that emerge when many molecules or atoms interact. One such emergent property is crystalline order among atoms; another is its opposite, the glass state, which involves rapid cooling through a melting temperature.

'Frustration', the inability of a system to find a unique ground state, is a common feature of disordered solids. Frustration can also arise in an ordered solid solely from geometric considerations, and this is increasingly being recognized as an organizing principle that governs a wide range of physical phenomena in the collective behaviour of atoms. Yet, contrary to the implication of their name, frustrated systems actually have more configurational options than conventional collective systems. The basic concepts of frustrated systems are turning out to have practical uses in areas from microelectronics to drug delivery.

The full reality of a solid is determined not only by the positions of its atoms or molecules, but by their intrinsic local properties — their magnetic, electric or rotational degrees of freedom, or moments. A ferromagnetic crystal, for instance, has two types of order: that of its atoms, and that defined by the mutual alignment of the atomic magnetic moments. When the local-moment order has a lower energy than the crystalline order, the crystalline lattice imposes its symmetry on the local moments — for instance, the magnetic field inside the ferromagnet has the same periodicity as the crystal lattice. Alternatively, the local-moment order can possess its own symmetry, dictated by the interactions between moments. For the ferromagnet, this symmetry is simply the mutual alignment of magnetic moments, which is compatible with any crystal symmetry.

What happens when the interaction between moments is not compatible with the crystal symmetry? Take antiferromagnetism, which occurs when the magnetic moments

of neighbouring atoms prefer anti-alignment. Although such an interaction is compatible with a square lattice, when such moments reside on a triangular lattice there is no simple ordered pattern that satisfies all of the bonds. In this situation, the unsatisfied bond is said to be frustrated. In a system composed of many such triangles, the threefold multiplicity of lowest and equal energy states translates into a large amount of low-energy entropy — geometric frustration.

Qualitatively new types of property can emerge when such systems are cooled to temperatures well below the energy of moment–moment interactions. An example is the 'spin-liquid' state in a material ($\text{SrCr}_2\text{Ga}_2\text{O}_{10}$) that consists of corner-shared triangles. Unlike the low-temperature state of an antiferromagnet, the spins in a spin-liquid behave like water molecules in the liquid state — when looked at in pairs, there is a preferred mutual arrangement of molecules (or spins), but this arrangement does not extend beyond the nearest neighbour.

Although geometric frustration is being recognized as a new way to classify magnets in naturally occurring minerals such as spinels and pyrochlores, the most common example of the phenomenon is found in ordinary water ice — or 'ice-one'. (There are many different crystal phases of ice, accessible under applied pressure. There is even an ice-nine, although it doesn't possess the properties proposed by Kurt Vonnegut.) Here, despite the well-ordered oxygen lattice, the hydrogen positions are completely disordered, down to the lowest accessible temperatures. This disorder was predicted by Linus Pauling in the 1930s on the basis of missing entropy in specific heat measurements, and has now been observed directly.

'Spin ice' is an analogy between the hydrogen positions in ice and the magnetic moment directions in pyrochlore-based magnets. Like the hydrogen positions in ice-one, there is complete disorder among the magnetic-moment directions, despite the fact that they occupy a well-ordered lattice. Unlike ice-one, however, the spin-ice lattice is formed at a temperature that is much higher than that at which the moments interact, which allows the study of ice-like dynamics over a broad range of energies. Thus, although it is of greater practical use, ice-one is the worst system for studying its own dynamics.

Geometric frustration might exist in other non-magnetic materials, such as the ceramic ZrW_2O_8 . This material exhibits negative thermal expansion: it expands when cooling, while retaining its cubic, symmetrical structure. This is in contrast to most other materials, which not only shrink when cooled, but

Geometric frustration

When magnetic moments that prefer anti-alignment reside on a triangular crystal lattice, there is no simple pattern that satisfies all of the bonds.

also adopt a lower-symmetry structure. Such behaviour is accompanied by a brief softening of the lattice, a process absent from ZrW_2O_8 .

Other structural versions of geometric frustration might exist in materials that possess both a body-centred unit cell and a local atomic tendency to exhibit ferroelectricity, which are mutually incompatible on the grounds of symmetry. Synthesizing such materials is a challenge, but carries potential commercial benefits, in the form of flash memory or smart cards with a variety of uses from toll booths to supermarkets.

When moment interactions fall significantly below the crystallization energy of the structure, as in the above examples, the structure dictates the behaviour at low energies through frustration. However, if the structure itself can respond to the frustration, unexpected patterns emerge. The blue phase of cholesteric liquid crystals is one example. Another is the design of self-assembled structures, such as nanotubules, for drug and gene delivery. Nanotubules are variants of liposomes, lipid electrolyte bilayers that curl up into a space-enclosing ball. When the polyelectrolytes and oppositely charged lipid membranes have very different charge densities, simple layered structures or balls are not sterically possible, leading to nanotubule formation.

Liposomes have micro-encapsulating and chemical properties that are valuable in textiles, cosmetics and drug delivery, in which specific sites in the body can, in principle, be targeted for administration of small amounts of drugs. Examples are retinoic acid for skin care, an influenza vaccine, and a topical anticancer cream. Fine-tuning the size and shape of nanotubules by understanding the shape-forming mechanisms of geometrical frustration should allow even greater flexibility for designing drugs.

Arthur P. Ramirez is at the Materials Integration Science Laboratory and the Institute for Complex Adaptive Matter, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

FURTHER READING

Ramirez, A. P. in *Handbook of Magnetic Materials* Vol. 13 (ed. Buschow, K. H. J.) 423–520 (Elsevier Science, Amsterdam, 2001).

Harris, M. J. *et al.* *Phys. Rev. Lett.* **79**, 2554–2557 (1997).
Wong, C. C. L. *et al.* *Science* **288**, 2035–2039 (2000).

W. WISNIEWSKI/EPIC PICTURE AGENCY/CORBIS



Ice provides the commonest case of frustration.

Now you see them, now you don't

David Wark

Evidence has been growing that the fundamental particles known as neutrinos oscillate — one type of neutrino can transform into another type. A well-placed experiment now points to the definitive answer.

Popular descriptions of the history of science are littered with stories of a much-loved theory swept aside by a single key experiment and replaced by a new, very different theory. In reality, it usually doesn't happen like that. Proponents of the existing theory will simply claim that any single experiment is wrong, or try to argue around it. Thus, acceptance relies on effects being seen in a number of experiments that address different aspects of the new and old models. This has certainly been true in neutrino physics, where the case for a radical overhaul of theory has been building for thirty years. Stunning new results from the Kamioka Liquid scintillator Anti-Neutrino Detector (KamLAND), announced at a conference last month and now formally published¹, have cut through a thicket of proposed new models, and show that the simplest and most elegant is nature's choice.

The existence of neutrinos was first proposed in 1930, to explain an anomaly seen in experiments on the radioactive β -decay of atomic nuclei. Neutrinos — fundamental particles similar to electrons, but much lighter and with no electric charge — interact only weakly with normal matter and can penetrate a star with negligible attenuation. This led Ray Davis and colleagues to seek the first direct experimental evidence for thermonuclear reactions inside the Sun, by looking for the neutrinos that these reactions produce. When their experiment in the Homestake Mine, South Dakota, measured fewer neutrinos than expected², explanations ranging from experimental error to flaws in solar modelling were proposed.

The most elegant suggestion came from Samoil Bilenky and Bruno Pontecorvo³. In the standard model of particle physics, there are three types (or 'flavours') of neutrino — called electron, muon and tau neutrinos — all with zero mass. Bilenky and Pontecorvo pointed out that if neutrinos did have mass, quantum mechanics would allow the electron neutrinos emitted in the core of the Sun to change into one of the other types of neutrino, and thereby evade detection by the Davis experiment (which was sensitive only to electron neutrinos). They called this phenomenon neutrino oscillations.

At first, this explanation was considered rather speculative, but subsequent solar-neutrino experiments ruled out experimental error or flaws in solar modelling,

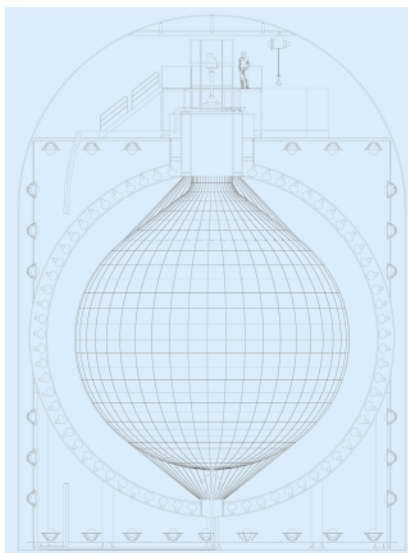


Figure 1 The KamLAND detector, in the Kamioka Mine, Japan. The transparent central balloon holds about 1 kilotonne of scintillating liquid. When an anti-neutrino interacts with protons in the liquid, the light emitted is picked up by the surrounding array of nearly 2,000 photomultiplier tubes, mounted in an 18-m-diameter sphere. The outer, water-filled volume, with more than 200 photomultiplier tubes, acts as a 'veto' for other particles entering the detector from the surrounding rock, especially muons.

Neutrino oscillations gained theoretical credibility when Stanislav Mikheyev and Alexei Smirnov⁴, building on the earlier work of Lincoln Wolfenstein⁵, showed that interactions with matter in the Sun could greatly enhance neutrino oscillations. Further support came from the Japanese Super-Kamiokande experiment, designed to pick up 'atmospheric' neutrinos coming from cosmic-ray interactions in the Earth's atmosphere⁶. The variations in the number of atmospheric neutrinos seen could not be explained within the standard model, but matched exactly the predictions for neutrino oscillations.

In 2001, Bilenky and Pontecorvo's explanation seemed to have been beautifully confirmed by results⁷ from the Sudbury Neutrino Observatory in Canada. In a single experiment, the number, or flux, of all flavours of neutrinos from the Sun was measured, and a separate measurement of

the electron-neutrino flux was also recorded. The total flux did indeed correspond to the number predicted by solar models, but the electron-neutrino flux was only about a third of the predicted value, showing that most of the electron neutrinos had oscillated into other flavours before reaching the Earth.

Or had they? The Super-Kamiokande results clearly established that electron neutrinos disappear, and the Sudbury results implied that they change flavour, but the exact mechanism was still open to question. Never underestimate the creativity of theorists — dozens of other models were proposed to describe the data, with mechanisms ranging from new properties of neutrinos, to the effects of higher dimensions, to violations of the basic assumptions of special relativity⁸. What was needed was an observation of oscillations using neutrinos from a known, well-calibrated source that would really test the key predictions of the model.

KamLAND⁹, an ingenious experiment in Japan, has taken advantage of some good luck to do just that. Like Super-Kamiokande, the experiment is built in an underground cavity in the Kamioka Mine (all these experiments must be located deep underground to screen out other high-energy particles from cosmic rays). KamLAND consists of a spherical volume of liquid scintillator, which emits light when hit by energetic charged particles, surrounded by a buffer liquid and water, which act as a shield against particles produced in radioactive decays in the surrounding rock (Fig. 1). The scintillator is watched by almost 2,000 photomultiplier tubes — light detectors so sensitive that they can detect a single photon.

KamLAND detects not electron neutrinos, but their anti-matter counterpart, electron anti-neutrinos. The anti-neutrinos interact with protons in the material of the detector to produce a neutron and an anti-electron, or positron. The positron immediately generates a tiny flash of light, but the neutron wanders around losing energy until it is captured in the detector material, producing a slightly delayed flash of light. These flashes are detected by the photomultiplier tubes, the second flash signalling that the first flash really was from an anti-neutrino interaction and not from some other, spurious interaction. The amount of light detected gives a measure of the anti-neutrino's energy and, using the timing of signals at

KAMLAND COLLABORATION

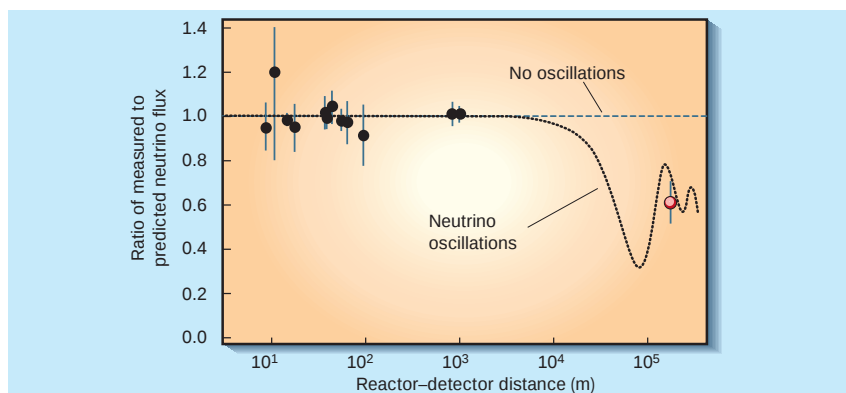


Figure 2 The right place at the right time. Earlier measurements (black points) of the ratio of measured to predicted flux of anti-neutrinos emitted from nuclear reactors have seen no trace of neutrino oscillations — the ratio is always one. But the first data from the KamLAND detector¹, ideally placed at the centre of a circle of reactors, give a lower ratio (red point), exactly as expected if the three types of neutrino can oscillate from one type to another.

the relative positions of the photomultiplier tubes, the location of the anti-neutrino interaction inside the detector can be reconstructed (which is another good way to discriminate against radioactive background signals from the edge of the detector).

None of this is luck — it is good experimental design. The luck is in the source of the anti-neutrinos: nuclear reactors, which are a prodigious source of electron anti-neutrinos and which have been exploited by several experiments around the world. Japan generates a large part of its electricity from nuclear power, and it so happens that KamLAND is surrounded by a number of powerful reactors, all at about the same distance — a lucky accident that means that the effects of oscillations (which vary with the distance the anti-neutrinos travel) will add up rather than average out between the different reactors. By a second lucky accident, this distance turns out to be sufficiently large that oscillations, at the level implied by the solar-neutrino results, should produce a noticeable reduction in the number of electron anti-neutrinos seen in KamLAND.

The KamLAND collaboration of US, Japanese and Chinese physicists has now announced the results from its first 145 days of data-taking¹. They have measured the flux of electron anti-neutrinos and compared it with the predicted value (which is known quite precisely from the power produced by the reactors) — the ratio is 0.61, significantly less than 1 (Fig. 2). What is exciting for neutrino physicists is that this level of suppression is exactly what was predicted, based on the results of solar neutrino experiments. It is, however, not the same level of suppression seen in the solar experiments, where the equivalent ratio for electron-neutrino flux is only 0.35. That is because the Earth, through which KamLAND's anti-neutrinos travel, is much less dense than the Sun and so causes less enhancement of the oscillations. But the two different levels of suppression corre-

spond to the same underlying physics — a coincidence that is true for neutrino oscillations, but not for any of the other mechanisms that theorists have speculated might be the cause of the observed atmospheric and solar neutrino deficits. KamLAND has therefore provided vital confirmation of the real reason for the neutrino disappearance and flavour change seen by the other experiments.

So what next? These results from KamLAND are based on a small sample of data. More data should reveal whether the energy spectrum of the anti-neutrinos is distorted, which is another clear signature of oscilla-

tions and would enable the properties of neutrinos to be determined more precisely. Meanwhile, other experiments are planned or are being built around the world to explore the phenomenon of neutrino oscillations in much greater detail. Hopefully, these efforts will culminate in the building of a 'neutrino factory', which would fire intense neutrino beams at detectors thousands of kilometres away. Such an experiment could show whether the laws that govern neutrino behaviour are different if time moves backwards instead of forwards — an effect that is being sought in all the fundamental laws of physics to explain why there is more matter than anti-matter in the Universe. Thus, the solution to one mystery brings the opportunity to search for the solution to another, and neutrino physics looks set to remain exciting for years to come. ■

David Wark is at the University of Sussex, and in the Particle Physics Department, Rutherford Appleton Laboratory, Chilton, Didcot OX11 0QX, UK.
e-mail: d.l.wark@rl.ac.uk

1. XXI Texas Symposium on Relativistic Astrophysics, Florence, 9–13 December 2002; Eguchi, K. *et al. Phys. Rev. Lett.* **90**, 021802 (2003).
2. Cleveland, B. T. *et al. Astrophys. J.* **496**, 505–526 (1998).
3. Bilenky, S. M. & Pontecorvo, B. JINR-E2-10545, reprinted in *Nucl. Part. Phys.* **7**, 149–152 (1977).
4. Mikhayev, S. P. & Smirnov, A. Yu. *Sov. J. Nucl. Phys.* **42**, 913–917 (1985).
5. Wolfenstein, L. *Phys. Rev. D* **17**, 2369–2374 (1978).
6. Fukuda, Y. *et al. Phys. Rev. Lett.* **81**, 1562–1567 (1998).
7. Ahmad, Q. R. *et al. Phys. Rev. Lett.* **89**, 011301 & 011302 (2001).
8. Gago, A. M. *et al. Phys. Rev. D* **65**, 073012 (2002).
9. Piepke, A. *Nucl. Phys. Proc. Suppl.* **91**, 99–104 (2001).

DNA repair

Damage alert

Jiri Bartek and Jiri Lukas

Radiation and other harmful influences frequently damage our genes, potentially causing diseases such as cancer. New work reveals a surprising mechanism that notifies the cellular defence system about DNA damage.

Our genetic blueprint is constantly assaulted by adverse environmental and cellular influences, such as ultraviolet or ionizing radiation and various chemicals¹. Fortunately, these massive attacks on our DNA are largely counterbalanced by promptly deployed, multifaceted surveillance and rescue operations. Among the many types of damage that arise in DNA, the most deadly are double-strand DNA breaks (DSBs), where both phosphate backbones of the double helix are severed. If left unrepaired, such damage can result in chromosomal defects and lead to inborn diseases or cancer¹. On page 499 of this issue, Bakkenist and Kastan² provide eagerly awaited insight into how the vital response to DSBs is initiated in human cells.

Researchers have known for years that the cellular response to DSBs requires the protein kinase ATM — the product of the *ATM* gene,

which is mutated in the human disease ataxia-telangiectasia^{3,4}. People with this disease suffer immune deficiency, partial brain degeneration, sterility, sensitivity to radiation and a higher risk of cancer, and their cells fail to respond to DSBs. In contrast, normal cells that are victim to radiation-induced DSBs can activate ATM, which then adds phosphate groups to (phosphorylates) several target proteins. These proteins are in turn linked into a signalling network that slows the cells' progression through the cell-division cycle, and stimulates DNA repair. If the DNA damage is irreparable, the network eliminates affected cells by a suicide programme^{3–7}.

But what tells ATM when to start phosphorylating its protein targets? How do the 'dormant' and activated forms of ATM differ? And how does this mechanism ensure that proteins throughout the cell nucleus

respond in a coordinated way, given that damage is often limited to a few DSBs within three billion base pairs of DNA? Bakkenist and Kastan² address all of these questions.

The authors first show that, in the absence of DNA damage, ATM proteins are locked together in pairs in a tight embrace that prevents them from forming any promiscuous liaisons with other proteins (Fig. 1). Following DSB formation, the two ATM proteins divorce, their separation being triggered when they phosphorylate each other at a particular amino acid (the serine at position 1981 in the protein). The newly single ATM proteins then target their cellular substrates, setting off the train of events that leads to either DNA repair or cell death. Given that ATM is a giant protein, identifying this one amino acid was like finding a needle in a haystack, and is a significant achievement.

This response to DNA damage has some interesting features. First, there are several ways in which cells keep other kinases inactive, but the mechanism now described for ATM is new. Second, most other cellular responses to DNA damage operate only in actively dividing cells^{1,3–6}; this is both because errors often creep in during DNA replication, and because it is especially important to repair damage then so as to prevent mistakes being passed on to newly generated cells. By contrast, the ATM response works in resting, non-proliferating cells as well^{2,8}. Finally, the sensitivity, extent and speed of the ATM response are truly astonishing. Doses of irradiation that cause only a few DSBs in a human cell activate the majority of ATM within minutes². And induction of just two DSBs per cell is enough to induce the crucial 'autophosphorylation' of ATM.

This suggested to Bakkenist and Kastan that the DSBs must cause changes in the nucleus that can switch on ATM even at a distance from the actual DNA breaks. Changes in chromatin — a complex structure formed by chromosomal DNA and a family of nuclear proteins called histones⁹ — seemed to fit the bill, and indeed the authors provide evidence that chromatin restructuring is involved. Specifically, drugs that do not damage DNA, but cause the normally tight-knit chromatin to loosen up, also activate ATM and lead to ATM-mediated phosphorylation of one protein target, p53, throughout the nucleus. (This protein directs cells towards either delaying their division or dying.) ATM substrates that normally accumulate at and repair DSBs remain unphosphorylated. This might be expected given that no breaks have occurred. But it also hints that further events or 'adaptor' proteins may be needed to direct activated ATM to such substrates after DNA damage.

It seems plausible, then, that DSBs induce changes in chromatin that lead to ATM activation in human cells. This fascinating new concept is consistent with observations that, in

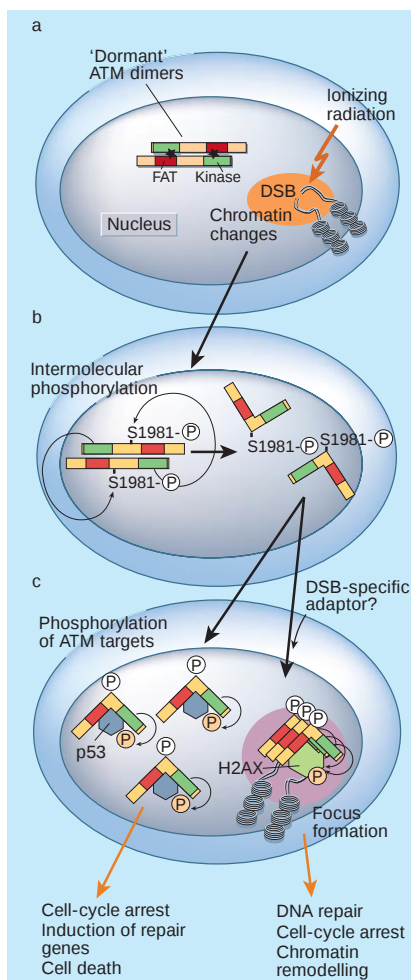


Figure 1 Arming cells to respond to DNA damage.

a. As Bakkenist and Kastan² show, in unstressed cells the ATM kinase forms dormant dimers, distributed throughout the nucleus. The 'FAT' domain of one ATM unit interacts with the enzymatic (kinase) domain of the other, and this probably locks ATM into a state in which it cannot interact with and phosphorylate its protein targets. Generation of DNA double-strand breaks (DSBs) by, for instance, ionizing radiation causes a local disruption of chromatin structure. This causes intermolecular modifications within ATM dimers that ultimately lead to their activation. **b.** The two bound ATM proteins phosphorylate each other on serine residue 1981. This triggers their dissociation and allows them to recognize their targets. **c.** Each free ATM can phosphorylate several nuclear proteins, most of which (such as histone H2AX) accumulate at sites of DNA damage and form 'foci'. The tumour suppressor p53 is also an ATM substrate, targeted by ATM throughout the nucleus. Ultimately, the cell repairs its damaged DNA; if this is not possible, the cell is eliminated.

yeast, mutations in genes coding for components of the so-called RSC chromatin-remodelling complex result in dramatically decreased tolerance to DNA damage¹⁰. Perhaps the mutant yeast suffer because they

cannot generate crucial chromatin changes in response to DNA breaks. The precise mechanism might be different in yeast: the essential serine is found in ATM in higher eukaryotes such as mice, frogs and humans², but not in related kinases such as MEC1 and TEL1 in yeast⁶, a lower eukaryote. So the autophosphorylation of dimeric ATM could be an evolutionarily younger mechanism, perhaps like the addition of p53 to the damage-response arsenal⁷. But the fact that yeast RSC mutants and MEC1 mutants show similar sensitivity to DNA damage hints that all eukaryotes might activate the response to DSBs through chromatin changes and ATM-related kinases.

This excellent paper² raises a host of intriguing questions. One is whether ATM senses the physical changes in 'relaxed' chromatin directly, or whether additional sensors or mediators are involved, such as DSB-interacting complexes or factors that regulate chromatin structure^{5,9–11}. In addition, the story of how ATM recognizes its targets remains unfinished. p53 is phosphorylated by ATM outside damaged regions, apparently without needing other proteins. But most targets are recognized by activated ATM only at DSBs, and may well require adaptor proteins so that they connect with ATM there but not elsewhere.

And there are other problems to be solved. Following DNA damage, how do the individual proteins within an ATM dimer stop inhibiting each other even though they are still in contact? And how is the inactive state restored once DNA is repaired? Do cells destroy the active ATM and make fresh, inactive proteins? Or is the activating phosphate stripped from ATM and the protein made dormant once more?

Another question is, what is so specific about the chromatin changes associated with DSBs? Chromatin remodelling occurs frequently as genes are expressed, and during every cell-division cycle when DNA is replicated, yet ATM seems to ignore such changes. Is there something unique about the DSB-evoked chromatin alterations? Or is there a threshold level of change, reached only with DNA damage or when DNA replication goes wrong? That remains to be seen. In yeast, there is indeed a threshold for activation of the DNA-damage response during DNA synthesis¹². But, given Bakkenist and Kastan's finding² that very low levels of DSBs activate ATM, if such a threshold exists in human cells it is likely to be controlled at later stages. All of this provides fertile ground for future research.

Jiri Bartek and Jiri Lukas are in the Department of Cell Cycle and Cancer, Institute of Cancer Biology, Danish Cancer Society, Strandboulevarden 49, DK-2100 Copenhagen, Denmark.
e-mail: bartek@biobase.dk

1. Hoeijmakers, J. H. J. *Nature* **411**, 366–374 (2001).

2. Bakkenist, C. J. & Kastan, M. B. *Nature* **421**, 499–506 (2003).

3. Shiloh, Y. *Curr. Opin. Genet. Dev.* **11**, 71–77 (2001).
4. Kastan, M. B. & Lim, D.-S. *Nature Rev. Mol. Cell Biol.* **1**, 179–186 (2000).
5. Abraham, R. T. *Genes Dev.* **15**, 2177–2196 (2001).
6. Zhou, B.-B. S. & Elledge, S. J. *Nature* **408**, 433–439 (2000).
7. Wahl, G. M. & Carr, A. M. *Nature Cell Biol.* **3**, 277–286 (2001).
8. Lukas, C. *et al. Cancer Res.* **61**, 4990–4993 (2001).
9. Wolffe, A. P. & Guschin, D. J. *Struct. Biol.* **129**, 102–122 (2000).
10. Koyama, H., Itoh, M., Miyahara, K. & Tsuchiya, E. *FEBS Lett.* **531**, 215–221 (2002).
11. Rouse, J. & Jackson, S. P. *Science* **297**, 547–551 (2002).
12. Shimada, K., Pasero, P. & Gasser, S. M. *Genes Dev.* **16**, 3236–3252 (2002).

Astronomy

Distant planet is the hottest yet

Timothy M. Brown

The first planet beyond our Solar System to be detected by means of the transit method has now been found to orbit its star almost twenty times closer than Mercury orbits the Sun.

In recent years, more than 100 planet-like bodies have been detected orbiting distant stars. Until now, all of these 'extrasolar' planets have been located through their gravitational effect on their parent star. An alternative method of planet-finding is to look for 'transits' — the passage of a planet in front of its star that causes a dip in the light intensity seen at the Earth. Building on brightness observations by Udalski *et al.*^{1,2}, Maciej Konacki and colleagues³ (page 507 of this issue) present delicate measurements that may confirm the first detection of an extrasolar planet by the transit method. This discovery is important for two reasons. First, the validation of the transit method has implications for many other planet-detection projects. Second, the inferred planet is closer to its star than any other planet yet seen, and further study may help to clarify some mysteries of the planet-formation process.

Since the first discovery⁴ of a planetary companion to a Sun-like star, in 1995, extrasolar planets have confounded theoretical expectations. A substantial fraction of them are similar in mass to Jupiter, but move in orbits very close to their stars. The sizes of planetary orbits are usually given in terms of the astronomical unit (AU), which is defined as the average distance between the Earth and the Sun. The closest planet to our Sun is Mercury, which orbits at a distance of about 0.4 AU. In contrast, about a dozen extrasolar planets have been found with orbital radii between 0.05 and 0.1 AU. How such massive planets arrived in such tiny orbits remains a difficult question to answer, although there are some plausible theories^{5,6}.

In general, it should be much easier to detect a dip in a star's brightness resulting from the transit of a planet than to detect the gravitational tug of the planet on the star. Furthermore, using modern imaging detectors, thousands of stars can be monitored simultaneously. Yet, of the 100 or so extrasolar planets found, this is the first to be detected through its transit (one transiting extrasolar planet has been seen, but it was

discovered through its gravitational pull on its star^{7,8}). This is because only a small fraction of extrasolar planets actually produce transits. Transits occur only if the Earth happens to lie nearly in the plane of the planet's orbit; otherwise, the planet never passes across the disk of its star, as seen from Earth (Fig. 1). In fact, the chance of producing transits increases with decreasing orbital radius, so close-in planets are much more likely to be seen in transit than are those that orbit far from their stars. Hence, the transit method is particularly suited to studying planets in small orbits. Also, the depth of the brightness dip provides a direct measurement of the relative sizes of the planet and the parent star.

The past three years have seen vigorous attempts to locate transiting planets by searching for their characteristic brightness dips (Fig. 1). The first lesson learnt from these efforts has been that there are many double- and multiple-star systems that can masquerade as transiting planets, and the trick is to separate the real planets from these interlopers. The success of Konacki *et al.*³ is largely attributable to a systematic method for doing this. They began with brightness

records for 59 stars that seemed to display a transit-like dip, put forward from a larger sample analysed by Udalski *et al.*^{1,2}. Through a series of tests and observations, Konacki *et al.* were able to reject about 90% of these candidates, judging them likely to be stellar, not planetary, systems. The final step was to examine the spectra of the remaining half-dozen stars. The light from the stars suffers a 'Doppler shift', with its wavelength being lengthened (or shortened) if the star is moving away from (or towards) the observer. The gravitational pull of an orbiting low-mass planet causes a very slight variation in the star's radial velocity, which can be picked up in the Doppler-shifted spectrum of light from the star and can also be used to deduce the mass of the orbiting planet.

The measurements made by Konacki *et al.* were particularly difficult, because the stars in question are typically several hundred times fainter than those normally targeted using this 'radial-velocity' technique. Nevertheless, they found that the radial-velocity measurements for one of the surviving candidates, OGLE-TR-56, match the expectations for a planet that has a mass slightly lower than Jupiter's, but an orbital radius of only 0.023 AU, or just under 3.5 million kilometres. If this is indeed a planet, it has the smallest orbital radius yet seen (by a factor of about two) and is consequently the hottest planet known — its dayside temperature is probably about 1,900 K. Konacki *et al.* suggest that the new planet may have a different history from other known planets, in that it may have lost a significant amount of mass to its parent star before settling into its current orbit.

Exciting as these data are, the sampling is quite poor over the full duration of the planet's orbit. Although non-planetary explanations of the observed variations seem to be ruled out, the planet interpretation will be stronger when the radial-velocity curve has been filled out with additional observations (see Fig. 1a on page 507). Planned space

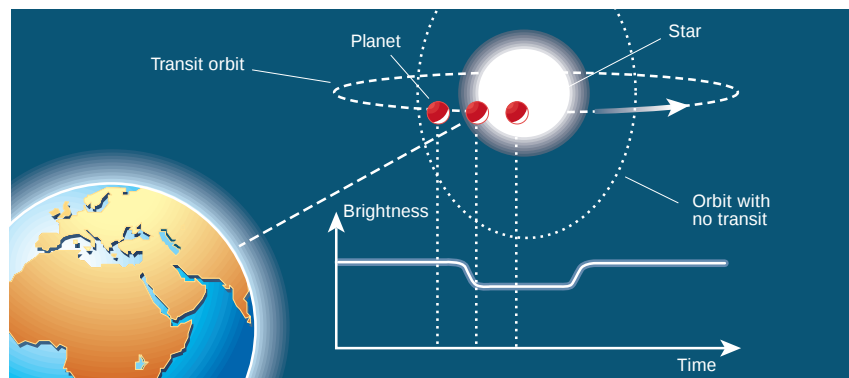


Figure 1 Planetary transit. The passage of a planet in front of a distant star produces a characteristic dip in the star's brightness, as perceived at the Earth. The apparent discovery of an extrasolar planet using this transit technique may now be confirmed by Konacki and colleagues' observations³. Although more than 100 extrasolar planets have been found, this is the first such planet to be detected through its transit: for the brightness dip to be seen, the line of sight from the Earth to the star must lie close to the plane of the planet's orbit.

experiments, such as the European Space Agency's Eddington and NASA's Kepler⁹ missions, will also search for extrasolar planets through their transit signatures. Avoiding the data deterioration caused by the Earth's atmosphere, these aim to locate planets as small as, or smaller than, the Earth. The success of Konacki *et al.*³ should inspire even greater enthusiasm for the promising projects soon to come.

Timothy M. Brown is at the High Altitude Observatory, National Center for Atmospheric Research, 3450 Mitchell Lane,

Boulder, Colorado 80303, USA.

e-mail: timbrown@hao.ucar.edu

1. Udalski, A. *et al.* *Acta Astron.* **52**, 1–37 (2002).
2. Udalski, A. *et al.* *Acta Astron.* **52**, 115–128 (2002).
3. Konacki, M., Torres, G., Jha, S. & Sasselov, D. D. *Nature* **421**, 507–509 (2003).
4. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
5. Goldreich, P. & Tremaine, S. *Astrophys. J.* **241**, 425–441 (1980).
6. Murray, N., Hansen, B., Holman, M. & Tremaine, S. *Science* **279**, 69–72 (1998).
7. Henry, G., Marcy, G., Butler, R. & Vogt, S. *Astrophys. J.* **529**, L41–L44 (2000).
8. Charbonneau, D., Brown, T., Latham, D. & Mayor, M. *Astrophys. J.* **529**, L45–L48 (2000).
9. Lissauer, J. *Nature* **419**, 355–358 (2002).

Biodiversity

The threat of small households

Nico Keilman

Many studies have suggested that the increasing global human population is having a negative effect on biodiversity. According to new work, another threat comes from the rising number of households.

Households in many countries have become smaller in recent decades. Between 1970 and 2000, the average number of occupants in households in less developed countries fell from 5.1 to 4.4. And in more developed nations, the decrease was from 3.2 to 2.5 people per household over the same period (the decline began earlier; Fig. 1). From their analysis of household dynamics in biodiversity 'hotspot' areas, Liu and colleagues¹ now argue (page 530 of this issue) that the decline in household sizes has unintended negative effects. The global human population has risen, not fallen, so smaller households means more households — and a higher demand for natural resources. This is in addition to the increased demand resulting purely from population growth.

Even before the writings of Thomas Malthus in the late eighteenth century, the balance between population and natural resources was a recurrent theme. Since ancient times, statesmen and philosophers have expressed opinions about such issues as the optimum number of people and the disadvantages of excessive population growth². Although some theorists see population expansion in a positive light^{3,4}, there is increasing concern about the negative consequences for resources⁵. Other things being equal, a larger population implies a greater demand for food, water, arable land, energy, building materials, transport and so on — a link that was first quantified some 30 years ago⁶. A population's age structure also influences economic growth and hence resource use: a rapid growth of the young age segments decelerates economic growth⁷.

More recently, scholars have acknowledged that another demographic variable — the number of households — also has an important role in resource consumption^{8–11}.

Even when the size of a population remains constant, more households imply a larger demand for resources. Household members share space, home furnishings, transportation and energy, leading to significant economies of scale. For instance, two-person households in the United States in 1993–94 used 17% less energy per person than one-person households¹¹.

To appreciate the different effects of population size and number of households on resource consumption on a larger regional scale, consider the following example⁸. In more developed regions, energy consumption increased by 2.1% per year over the period 1970–90. Population growth can explain 0.7 percentage points of this growth in energy usage, while changes in per capita energy use explain the remaining 1.4 points. However, an alternative analysis decomposes the growth in energy consumption into a factor that describes the growth in number of households and a factor describing per household energy use. This analysis shows that the household growth factor explains

1.6 percentage points of the energy-consumption increase — more than twice as much as the population growth factor.

Liu and colleagues¹ now draw our attention to household dynamics in biodiversity hotspot areas — regions that are rich in endemic species and threatened by human activities. They find that, during the years 1985–2000, the number of households in 76 hotspot countries increased by 3.1% per year, substantially faster than did the population (1.8% per year). So, average household size fell by about 1.3% per year. These changes relate to the group of 76 countries as a whole. For individual hotspot countries, more than 80% showed a pattern of greater growth in household numbers than in population. In 65 non-hotspot countries, however, population increased at roughly the same tempo as household numbers during 1985–2000.

Many of the world's most populated countries are hotspot countries (such as China, India, Indonesia, Brazil and Bangladesh). And most of the hotspot countries studied by Liu *et al.* (65 out of 76) belong to the group of less developed nations. We know that falling birth rates were an important driving force behind reductions in average household size in less developed countries in the 1990s (ref. 12). Despite these falling birth rates, however, the population in such countries did increase (because of decreased death rates, for instance). All of this might explain why increases in the number of households were relatively pronounced in hotspot countries¹.

Liu *et al.* also refer to projections of population size and the number of households over the next 15 years. These projections suggest that the divergence in population growth and household numbers will become more pronounced. So, the authors argue, it is crucial to consider average household size when assessing threats to biodiversity. Quantifying the impact of falling household sizes, and increasing household numbers, on biodiversity changes should have high research priority.

Small households have adverse effects on resource consumption both because they are less energy-efficient in themselves and

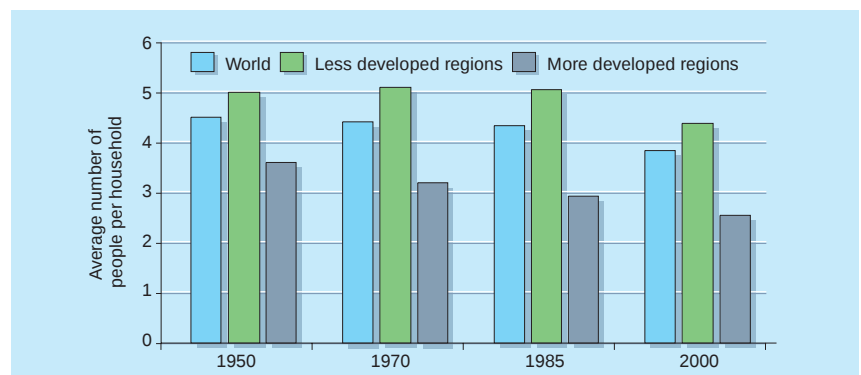


Figure 1 **Decline and fall in household sizes.** Data for 1950 and 1970 are taken from ref. 8; data for 1985 and 2000 are from ref. 17.

because they often reflect an increase in the number of households. If this increase could be stabilized at roughly the same level as population growth, the adverse effects might also stabilize. But could this be achieved? That depends on possible explanations for why household sizes have fallen in the first place. Some of these explanations are as follows. First, all other factors remaining the same, falling birth rates reduce population size, but do not affect the number of households; hence, household size is reduced. Second, increased material standards of living have an effect. Extended households are observed in countries in an early stage of development¹³. When these countries attain a higher standard of living, some institutions — such as social-security systems — provide the assurance against risks that were formerly supplied by the extended household.

Third, social, economic and cultural theories of demographic behaviour point to a variety of reasons why individuals prefer to live in small households^{14–16}. These include less adherence to strict norms; less religiosity and increased individual freedom on ethical issues; female education, which has led to women having greater economic independence and also facilitates divorce; more assertiveness in favour of symmetrical gender roles; the contribution of women to the labour market; increased economic aspirations; and residential autonomy. Fourth, population ageing reduces household size. This is a direct consequence of two facts: increased longevity leads to longer periods of time when children do not live with their parents; and the greater mortality of men, together with

the usual age difference between spouses, results in many widows who live alone.

Smaller households, then, are the result of processes that cannot be reversed (such as modern contraception and liberalization from norms) or that we value for a number of reasons (such as women's emancipation). So policy interventions will have to focus on the average household resource consumption, in order to combat the adverse effects of smaller households. ■

Nico Keilman is in the Department of Economics, University of Oslo, Blindern, N-0317 Oslo, Norway. e-mail: n.w.keilman@econ.uio.no

1. Liu, J., Daily, G. C., Ehrlich, P. E. & Luck, G. W. *Nature* **421**, 530–533 (2003); advance online publication, 12 January 2003 (doi:10.1038/nature01359).
2. Cohen, J. *How Many People Can the Earth Support?* (Norton, New York, 1995).
3. Boserup, E. *Population and Technological Change* (Univ. Chicago Press, 1981).
4. Simon, J. *The Ultimate Resource 2* (Princeton Univ. Press, 1996).
5. *Population, Environment and Development: The Concise Report* (United Nations, New York, 2001). <http://www.un.org/esa/population/publications/concise2001/C2001English.pdf>
6. Ehrlich, P. & Holden, J. *Science* **171**, 1212–1217 (1971).
7. Crenshaw, E., Ameen, A. & Christenson, M. *Am. Soc. Rev.* **62**, 974–984 (1997).
8. MacKellar, F. L., Lutz, W., Prinz, C. & Goujon, A. *Pop. Dev. Rev.* **21**, 849–865 (1995).
9. Cramer, J. *Demography* **35**, 45–65 (1998).
10. Jiang, L. *Population and Sustainable Development in China* (Thela Thesis, Amsterdam, 1999).
11. O'Neill, B. & Chen, B. *Pop. Dev. Rev. (Suppl.)* **28**, 53–88 (2002).
12. Bongaarts, J. *Pop. Stud.* **55**, 263–279 (2001).
13. Goode, W. *World Revolution and Family Patterns* (Free Press, New York, 1963).
14. Van de Kaa, D. *Pop. Bull.* **42**, 1–59 (1987).
15. Lesthaeghe, R. in *Gender and Family Change in Industrial Countries* (eds Mason, K. & Jensen, A.) 17–62 (Clarendon, Oxford, 1995).
16. Verdon, M. *Rethinking Households* (Routledge, London, 1998).
17. United Nations Centre for Human Settlements (Habitat) *Cities in a Globalizing World* (Earthscan, London, 2001).

particles were effectively hard spheres, with no Coulomb repulsion. Then, entropy — generally considered to be a measure of disorder — favours the f.c.c. crystal, as the number of configurations available to a particle localized about a lattice site in the f.c.c. crystal exceeds those accessible in a disordered fluid or the b.c.c. crystal³.

Although hard-sphere behaviour of polymer-based colloids could be achieved in model systems, there was a drawback: those colloids were opaque at even moderate densities, so little could be learned about their structure from light scattering. More transparent dispersions were sought, such as silica spheres coated with short hydrocarbon chains in a nonpolar solvent that eliminates surface charge⁴. In the 1980s, these organophilic silicas and the aqueous lattices sufficed for many studies of fluid-to-crystal transitions and other colloidal phenomena. But small silicas could not easily be made highly uniform in size and there can be extra, van der Waals attractions between the larger ones, so better colloidal hard spheres were sought. Eventually a standard emerged: poly(methylmethacrylate) (PMMA) spheres coated with a low-molecular-weight polymer⁵.

In a solvent that also contains soluble polymer, neighbouring spheres are pushed together by osmotic pressure due to expulsion of polymer chains from small gaps between the particles. This attractive force increases roughly linearly with polymer concentration and can easily cause a dilute gas-like dispersion to condense into a colloidal fluid, and then into a solid f.c.c. crystal. In reality, the hard spheres pass through an intermediate, random hexagonal close-packed (r.h.c.p.)⁶ phase and only slowly convert to the f.c.c. structure. For larger colloids or smaller polymer chains, the transition directly from 'gas' to f.c.c. crystal is more favourable⁷.

Thus long-range attractions or repulsions yield condensed phases with low density and coordination number, such as dense fluid or b.c.c. crystal phases. Short-range repulsions and attractions produce denser f.c.c. crystals with higher coordination number. But crystals with lower coordination numbers than the b.c.c. phase or more complex structures have not been achieved with spheres of a single size. Yethiraj and van Blaaderen¹ confront this issue by devising a model system in which the forces between particles can be tuned, combining a soft repulsion with a long-range, anisotropic attraction.

The authors laced PMMA spheres (with radii between 1 and 2 μm) with fluorescent dye and dispersed them in an organic mixture whose refractive index and density were chosen to aid confocal imaging of the spheres. The solvent also preserves sufficient dielectric contrast for an applied electric field to induce strong dipole–dipole inter-

Condensed-matter physics

Tunable colloidal crystals

William B. Russel

Microscopic particles dispersed in a solvent — a colloidal dispersion — can be a useful model for phase transitions and crystal nucleation. A colloid that can be 'tuned' using an electric field is a valuable new tool.

Fundamental advances in colloid science often depend on physical models, which are made by dispersing carefully tailored particles, less than a micrometre in size, in pure aqueous or organic liquids. Such dispersions can be characterized by methods such as light scattering and confocal microscopy, and the physical and chemical interactions between the particles, responsible for intriguing phases such as colloidal crystals (which behave like atomic solids), can be precisely controlled. On page 513 of this issue, Yethiraj and van Blaaderen¹ describe a new model system that can be tuned with an electric field to display phase transitions and unexpected crystalline structures.

Colloidal crystals first attracted interest

in the 1960s. In studies of the light scattered from dilute dispersions, a transition was detected from a disordered fluid to an ordered body-centred-cubic (b.c.c.) crystal when the screened (or reduced) Coulomb repulsions between the colloidal particles extended to length scales greater than the lattice spacing². In fact, this transition can be controlled: adding a small amount of salt decreases the range of the repulsive force, because the salt dissociates into ions that enhance the screening. As a result, the volume fraction (or density) of particles at the transition increases, and a denser, face-centred-cubic (f.c.c.) crystal structure is favoured. Adding even more salt leads to 'hard-sphere' transitions — as though the

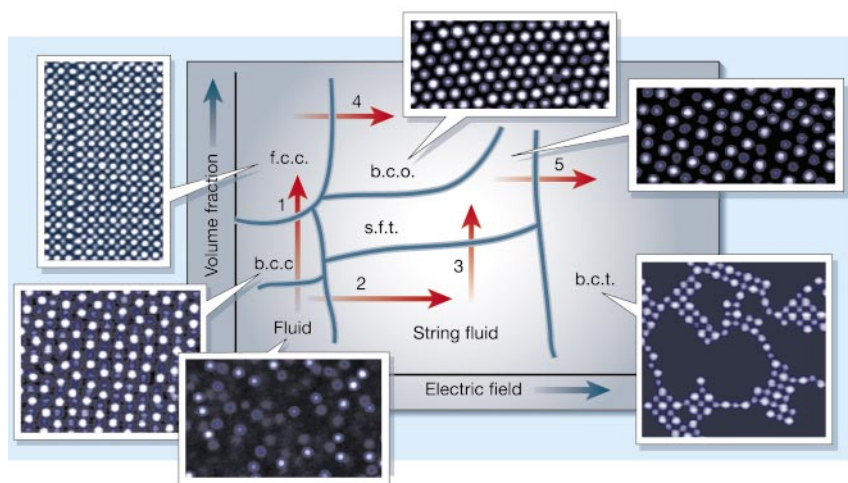


Figure 1 How to tune your colloidal crystal. Yethiraj and van Blaaderen¹ have devised a model colloid whose structure can be effectively 'tuned' by varying the applied electric field and the volume fraction, or density, of colloidal particles. At low field values, increasing the volume fraction (1) causes the random, fluid arrangement of colloidal particles to take on body-centred cubic (b.c.c.), then face-centred cubic (f.c.c.) crystal structures (shown in the insets). Increasing the electric field at fixed volume fraction (2), however, transforms the colloid into a string fluid structure. If the volume fraction is then increased (3), a space-filling tetragonal (s.f.t.) structure results. Alternatively, starting from high volume fraction with f.c.c. structure and increasing the electric field (4) produces the body-centred orthorhombic (b.c.o.) state. From the s.f.t. state, if the electric field is pushed up still further (5), the colloidal crystal takes on another, more open structure, known as body-centred tetragonal (b.c.t.).

actions, and the dye imparts charge and generates weak electrostatic repulsions that can be moderated by adding a soluble salt. So the interaction potential superposes a hard-sphere repulsion (considered to have infinite magnitude and zero range), an exponentially decaying repulsion (of finite magnitude, with its decay length dependent on salt concentration) and an orientation-dependent dipolar attraction/repulsion (with magnitude proportional to the electric field strength, and range comparable to the particle radius). The result is a three-dimensional phase diagram defined by the volume fraction, the range of the dipolar force and the strength of the electric field.

Yethiraj and van Blaaderen first sampled one plane of the phase diagram, looking at the effects of varying volume fraction and range of the electrostatic repulsion, in the absence of an electric field. They found fluid-to-b.c.c.-to-f.c.c. transitions for soft repulsions, fluid-to-f.c.c. transitions for harder electrostatic repulsions, and fluid-to-r.h.c.p. transition for nearly hard spheres, consistent with earlier results.

But then, looking at the variation with volume fraction and electric field, they found richer phase behaviour. At low values of electric-field strength, increasing the volume fraction induces a fluid-to-b.c.c.-to-f.c.c. sequence (Fig. 1). But maintaining a low volume fraction and increasing the field strength instead leads across a phase boundary from the isotropic fluid to a disordered phase in which the spheres condense into strings. Then, increasing the volume fraction at intermediate field strength drives a first-order transition, forming a space-filling

tetragonal (s.f.t.) crystal. More interesting still are the solid-solid transitions Yethiraj and van Blaaderen observed by increasing the field strength from the f.c.c. crystal to a body-centred orthorhombic (b.c.o.) structure, and then from the s.f.t. crystal to a body-centred tetragonal (b.c.t.) structure, the phase that

had been generally expected for hard spheres in a strong electric field. The tunability of the system through the electric field also means that the process can be reversed to observe complex melting phenomena, as one or more phase boundaries are traversed.

Yethiraj and van Blaaderen's model system, incorporating tunable dipolar attraction and soft repulsion, has revealed several new phase transitions for colloidal crystals. The newly identified crystalline phases, s.f.t. and b.c.o., as well as the b.c.t. phase expected in electric fields, represent anisotropic distortions of the b.c.c. lattice into more open structures with lower coordination numbers. Converting these fragile colloidal crystals into robust solids could have technological implications, and further tuning of the pair potential should reveal even more interesting crystal structures.

William B. Russel is in the Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544, USA.

e-mail: wbrussel@princeton.edu

1. Yethiraj, A. & van Blaaderen, A. *Nature* **421**, 513–517 (2003).
2. Monovoukas, Y. & Gast, A. P. *J. Colloid Interface Sci.* **128**, 533–548 (1989).
3. Hachisu, S. & Kobayashi, Y. *J. Colloid Interface Sci.* **46**, 470–476 (1974).
4. de Kruif, C. G., Jansen, J. W. & Vrij, A. in *Complex and Supramolecular Fluids* (eds Safran, S. A. & Clark, N. A.) 315–343 (Wiley-Interscience, New York, 1987).
5. Antl, L., Goodwin, J., Otterwill, R. & Waters, J. *Colloids Surf.* **17**, 67–78 (1986).
6. Pusey, P. N. & van Megen, W. in *Complex and Supramolecular Fluids* (eds Safran, S. A. & Clark, N. A.) 673–698 (Wiley-Interscience, New York, 1987).
7. Sperry, P. R. *J. Colloid Interface Sci.* **99**, 97–108 (1984).

Behavioural genetics

Family matters

David Haig

Differential activation of genes inherited from mothers and fathers will manifest itself as conflict in families. The effects are being explored experimentally with mice.

The evolution of maternal care involves a complex interplay between the genetic interests of mothers and those of offspring, with offspring predicted to favour higher levels of care than those favoured by mothers¹. This interaction is further complicated by the possibility of conflict within offspring genomes through a phenomenon known as genetic imprinting², and of conflicts within maternal genomes between genes that are inherited by a particular offspring and genes that are not³. A study by Hager and Johnstone (page 533 of this issue⁴) reveals some of the genetic complexities behind how these conflicts are resolved.

Imprinting is an important concept in this context. Two copies — alleles — of a gene are inherited by offspring, one from the father and one from the mother. In imprinting, some modification of the DNA sequence,

such as the addition of a methyl group to cytosine residues, means that one copy of the gene is inactivated and so only the paternally or maternally inherited version is expressed. Natural selection on genes expressed in offspring favours the solicitation of higher levels of maternal care when the gene is inherited from a father than from a mother because, in the former case, the allele has less 'interest' in the mother's future reproduction. This asymmetry sets up the possibility of a three-way conflict within offspring genomes — among unimprinted genes, maternally expressed imprinted genes (when the allele inherited by the offspring from its father is silent), and paternally expressed imprinted genes (when the allele inherited by the offspring from its mother is silent). For interpreting Hager and Johnstone's study, the important point is that the influence of

imprinted genes will show up as an effect of either the mother's or the father's genotype.

Hager and Johnstone compared litter sizes from crosses within and between two inbred strains of mice (CBA and C57/B6). Surprisingly, the father's strain had a highly significant effect on litter size, whereas the effect of the mother's strain was nonsignificant — that is, CBA males produced larger litters than did C57/B6 males when mated to females of either strain. So either a father's strain influences the number of eggs a female produces, or it influences the proportion of eggs that produce a newborn mouse. The authors suggest that an offspring's paternal genotype affects the proportion of embryos that are implanted in the mother's uterus but are subsequently resorbed. These possibilities are testable. Paternal effects on the number of eggs produced could be assessed by counting corpora lutea, the glandular bodies that develop from an ovarian follicle after its egg has been released. And differential survival could be estimated by counting embryos at different stages of pregnancy.

The genetic nature of the paternal influence on litter size is unclear. Hager and Johnstone's analysis tests for effects of the mother's strain and the father's strain, not for the effects of offspring genotype. Thus, CBA (mother) $\times$ C57/B6 (father) and C57/B6 (mother) $\times$ CBA (father) litters are treated as maximally distinct, because they differ from each other for both the mother's and the father's strain. But these litters have identical genotypes if parental sex is ignored with regard to strain, and they would be grouped together as a single class in most analyses testing for effects of offspring genotype.

Whether or not this is appropriate depends on the relative importance of imprinted and unimprinted genes. If litter size had been analysed with respect to the unimprinted genes carried by the offspring, several alternative explanations for the size difference could have been considered. For example, one could have tested for additive effects — was litter size influenced by the proportion of CBA alleles in offspring? Alternatively, tests could have looked for dominance effects — was litter size influenced by either the presence or absence of CBA or C57/B6 alleles in offspring?

Clearly, the possible paths of causation remain to be disentangled. Nevertheless, the authors' analysis emphatically rejects what, at first sight, would have been the most intuitive reason — that litter size is determined solely by the mother's strain. The influence of maternal genotype, if present, is weak, whereas there is strong evidence of the effects of offspring genotype. Further analysis is required to decide whether those effects are due to paternally expressed imprinted genes or unimprinted genes. But maternally expressed imprinted genes clearly do not have a major influence.

In Hager and Johnstone's experiment, newborn litters were cross-fostered to mothers of both inbred strains; no litters were raised by their actual birth mother. Hager and Johnstone assessed a female's level of milk supply by removing a foster mother from her 6-day-old pups for 4 hours, and then recording her weight loss during the 2 hours after their reunion. Not surprisingly, the foster mother's strain had a significant effect on the level of milk provisioning. Of greater interest is the fact that there was a notable interaction between a foster mother's strain and the maternal strain of her adoptive pups. Specifically, foster mothers lost more weight if they were of the same strain as their pups' birth mother. The absence of an equivalent interaction between the foster mother's and the father's strain implies that foster mothers can distinguish offspring from reciprocal crosses between CBA and C57/B6 mice. The nature of this parent-of-origin effect is unresolved, but there are several possible explanations. These include the effects of extranuclear genetic material passed on by the mother through the egg's cytoplasm; persisting maternal effects from having spent gestation in the uteri of birth mothers of different genotype; or the influence of maternally expressed imprinted genes.

Why should a female supply more milk to pups that carry maternally derived genes matching her own? Hager and Johnstone suggest two possible explanations — that pups may have evolved behaviours for extracting resources from foster mothers that are of the same strain as their own mother, or that mothers may have evolved mechanisms for preferentially provisioning their own offspring in communal nests shared with other females.

Hager and Johnstone's elegant experiments have uncovered intriguing interactions between the genotypes of mothers and offspring. These crosses, however, involved only two inbred strains of mice, and one cannot tell how far the results will apply to other strains, or to outbred mice. A further consequence of using inbred parents is that all offspring within litters have the same genotype — except that brothers and sisters inherit different sex chromosomes from their father — and every gene in a parent has an exact match in each offspring. So the experiments reveal little about competition within litters, or about the effects of maternal alleles that are not passed on to offspring. But Hager and Johnston have opened a promising line of investigation, and future experiments should help to unravel these complexities. ■

David Haig is in the Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA.
e-mail: dhaig@oeb.harvard.edu

1. Godfray, H. C. J. *Nature* **376**, 133–138 (1995).
2. Burt, A. & Trivers, R. *Proc. R. Soc. Lond. B* **265**, 2393–2397 (1998).
3. Haig, D. *Proc. Natl Acad. Sci. USA* **93**, 6547–6551 (1996).
4. Hager, R. & Johnstone, R. A. *Nature* **421**, 533–535 (2003).



100 YEARS AGO

The paper on electric automobiles read by Mr. H. F. Joel before the Institution of Civil Engineers on January 13 is one of great interest. The desirability of the automobile replacing horse traction from a sanitary point of view is probably admitted by everyone, and certainly the electric car would afford the best solution. Mr. Joel is of opinion that there is a great future before the electric automobile, which has already proved itself capable of running 100 miles on one charge and of performing much longer tours. This shows that even the storage battery of to-day is sufficiently good to give very satisfactory results; the author in his paper goes carefully into the results of the battery test made by the Automobile Club of France, and into the question of the ratio of weight of vehicle to weight of battery. Many valuable curves showing the relations between ton.mileage, total weight, useful load, &c., are given, and the paper is, on the whole, a valuable contribution to the subject.

From *Nature* 29 January 1903.

50 YEARS AGO

The granting of a Royal Charter to Queen Elizabeth College, University of London, marks the beginning of a new phase in the work of what has hitherto been known as King's College of Household and Social Science. Beginning as a department of King's College for Women in 1908, it gradually developed research and teaching in the scientific aspects of household and social work, until in 1920 there was introduced the first degree in these subjects — B.Sc. (Household and Social Science). By 1928, the department became an independent school in the University of London under the title of King's College of Household and Social Science. It played a prominent part in the development of dietetics as a specialized study, and began, in 1933, the first courses leading to a diploma in dietetics. Now, with its Royal Charter and its new name, Queen Elizabeth College (after Queen Elizabeth, the Queen Mother) [it] is breaking new ground. From October 1953 it will be training men and women undergraduates in the science of nutrition, leading to the new degree of B.Sc. (Nutrition). So far as we know, there is no other university in Britain which gives an undergraduate course for a first degree in this subject.

From *Nature* 31 January 1953.

Epidemiology

Conditions for dengue fever

Emerg. Infect. Dis. **9**, 86–89 (2003)

The Texan lifestyle could explain why dengue fever is so rare in the state, even though the infection is common just yards away across the border in Mexico. It's affordable air-conditioning that is responsible, Paul Reiter and colleagues suggest.

The dengue virus, spread by the *Aedes aegypti* mosquito, is endemic throughout South and Central America. Yet dengue is notably absent from the United States, even though *Ae. aegypti* is common in southern US states. Some experts have suggested that insect-control initiatives and climate differences could be the explanation.

Not so, say Reiter's team. By measuring antibodies to the virus in people's blood, they compared dengue infection rates in the Mexican city of Nuevo Laredo and in Laredo, Texas — essentially two halves of the same city straddling the Mexico–US border. On average, 16% of Nuevo Laredo residents had been exposed to dengue, compared with 1.3% in Laredo. The authors conclude that the single largest risk factor in Nuevo Laredo, accounting for 55% of dengue infections there, was the absence of air-conditioning.

Reiter *et al.* argue that air-conditioning — costing 4.5% of per capita gross domestic product in Laredo, compared with 24.2% in Nuevo Laredo — means that windows and doors are kept shut and people spend more time indoors.

Tom Clarke

Marine ecology

Fish that go to the cleaners

Curr. Biol. **13**, 64–67 (2003)

Many coral-reef fish seem to choose a base according to the grooming services available. Despite their small size and low numbers, then, cleaner fish that provide these services probably have a major influence on reef ecology.

Alexandra Grutter and her colleagues excluded the cleaner fish *Labroides dimidiatus* (see picture) from experimental sites on the Great Barrier Reef for 18 months. By the trial's end, these sites had half as many species and a quarter as many fish as those with cleaners. The effect on diversity was limited to mobile species that move between reefs, rather than those that set up home on one reef. Wandering fish are often large and predatory, and their presence or absence will affect the distribution and abundance of other species.

Cleaning on coral reefs is big business: an individual *L. dimidiatus* services on average more than 2,000 clients each day, removing



Little fish, large influence.

more than 1,000 parasites. Ungroomed fish can find their parasite load rising more than fourfold within 12 hours.

John Whitfield

Optical physics

Voltage manoeuvres on a microlens

Appl. Phys. Lett. **82**, 316–318 (2003)

A tiny lens with an electrically controllable shape and focal length has been developed by T. Krupenkin and colleagues. The microlens is simply a droplet of salty water, about 3 millimetres across and with a volume of 6 microlitres, sitting on a surface patterned with tiny electrodes. Applying a voltage between the electrodes and the conductive liquid changes the energy of the solid–liquid interface, causing the droplet to spread out or contract.

The edge of a spreading droplet tends to stick to the surface below, creating jerky rather than smooth motion and making the droplet's shape differ depending on whether it is advancing or retracting. To eliminate this, the researchers lubricated the interface by means of a thin film of an immiscible silicone oil. Voltages of up to 80 V produced a steady increase in the focal length of the lens-shaped droplet that was up to 20% of its initial value.

The researchers aim to extend the idea to make shape-changing waveguides and other components for photonic technologies.

Philip Ball

Biophysics

Solution to haemoglobin's structural paradox

Proc. Natl Acad. Sci. USA **100**, 517–520 (2003)

Haemoglobin was the first multi-subunit protein to have its atomic structure determined. Since then, more than 50 crystal structures for the molecule have revealed details of the changes it undergoes as it binds oxygen. Now Jonathan A. Lukin

et al. have determined how haemoglobin's four subunits fit together not in a crystal, but free in solution.

Previous structural studies have shown that, when it binds oxygen, haemoglobin shifts from a low-affinity 'T state' to a high-affinity 'R state', undergoing small movements of its amino acids and large-scale rotations of the subunits themselves. But these structures were derived from crystals, which could force the protein into unnatural conformations. Indeed, two R states have been identified, R and R2, one of which could be an intermediate in the conformational change.

In their approach, Lukin *et al.* obtained NMR spectra from haemoglobin solutions in which phospholipids (or bacterial viruses) were either present or absent. The spectra show small differences, because the presence of such liquid-crystalline lipids (or viruses) causes the orientations of haemoglobin to be biased slightly away from a fully random distribution. The differences reflect the relative alignments of individual chemical bonds and therefore of the complete protein subunits.

Lukin *et al.* found that the spectra fall halfway between those of the R and R2 states, and that haemoglobin in solution is highly dynamic, constantly shifting through a range of conformations, of which R and R2 are merely crystallized examples.

Christopher Surridge

Astronomy

Star formation out at the edges

Astrophys. J. **580**, L121–L124 (2002)

It had been thought that stars formed only in the gaseous parts of galaxies, where they could sometimes be traced by the glowing nebulae powered by massive stars. But new observations from a collaboration led by Ortwin Gerhard and Sadanori Okamura reveal that stars may also form outside these gas-rich regions.

Using the Subaru Telescope on Mauna Kea, Hawaii, and the European Southern Observatory's Very Large Telescope (VLT) on Paranal, Chile, Okamura's and Gerhard's teams turned their attention to the Virgo galaxy cluster, 50 million light years away.

Galaxy clusters are made up of hundreds of galaxies, but they also contain a vast amount of 'intracluster space' — the gaps between galaxies. By combining images and spectra from the two telescopes, the teams identified a sparse population of stars forming at the boundary between the halo of a galaxy and intracluster space.

This kind of 'isolated' star formation could be one explanation for the presence of stars in the haloes of galaxies — far from known star-forming regions.

Tom Clarke

Sensing temperature without ion channels

A remarkable gel under the shark's skin enables it to locate prey-rich thermal fronts.

Mammals use cold-sensitive ion channels to translate information about the temperature of their surroundings into electrical signals that are taken up by thermoreceptor nerve cells^{1,2}. Here I investigate the thermoelectric properties of an extracellular gel removed from the electro-sensors of sharks, and show that it develops significant voltages in response to tiny temperature gradients. This bulk property of the gel indicates that temperature can be translated into electrical information without the need for ion channels, a sensitivity that may help sharks to locate thermal fronts as feeding areas.

The ampullae of Lorenzini serve as acute electro-sensors for sharks³ and their relatives, as well as for a few other aquatic predators⁴, but they are also extremely sensitive to abrupt changes in temperature^{5,6}. Gel-filled canals connect each innervated ampulla to pores in the shark's skin. The clear, glycoprotein-based gel has electrical characteristics that are roughly similar to those of a semiconductor⁷, making it a likely medium for strong thermoelectric effects⁸.

Thermopower denotes the strength of the electric-field gradient that is generated by an applied temperature gradient across

a material. I measured the thermopower (also called the Seebeck coefficient, S) for gel samples (2 ml in a quartz container; Fig. 1a) collected from a black-tip reef shark (*Carcharinus melanopterus*) and a white shark (*Carcharodon carcharias*). A temperature gradient was induced by briefly heating one end of the sample holder, and was measured by using platinum thermometers; the resulting change in voltage was monitored by braided silver electrodes (Fig. 1a).

The system then returned slowly to thermal equilibrium (Fig. 1b). The constant voltage offset (usually between -8 mV and $+8$ mV) was presumably due to the resistive gel-electrode interface. Thermopower values were derived from the slope of data collected during heating by standard procedures⁹. Averaging over ten data sets from two samples for each species, thermopower values at room temperature were 370 ± 60 μ V per $^{\circ}$ C for the black-tip reef-shark gel and 240 ± 50 μ V per $^{\circ}$ C for the white-shark gel.

These values indicate that the gel is translating an abrupt thermal stimulus into an electric stimulus in the ampullae. The temperature sensitivity of ampullae (Fig. 1c) excised from dogfish (*Scyliorhinus canicula*) is of the correct order of magnitude⁶ (Fig. 1d)

for it to be explained by a thermoelectric mechanism. Assuming a gradient of 0.1 $^{\circ}$ C across the canal, the S values measured here would give rise to a voltage of roughly 30 μ V within the gel (a drop of 15 μ V in the warmer ampulla, and a rise of 15 μ V rise near the cooler pore). Extrapolating from measurements on the ampullae in the thornback ray (*Platyrhinoidis triseriata*), a drop of 15 μ V leads to a roughly 300% increase in the firing rate in the associated primary afferent¹⁰, which is consistent with the observed jump for dogfish ampullae.

In a thermopower mechanism, the transient nature of the dynamic temperature response arises from the natural progression to thermal equilibrium. Coupling the thermopower averages (about 300 μ V per $^{\circ}$ C) with the electro-sensory threshold established in behavioural experiments (about 0.05 μ V; ref. 11), the ampullae may even respond to changes of less than 0.001 $^{\circ}$ C. It is not known whether sharks rely on acute temperature sensitivity, but both white and basking sharks (*Cetorhinus maximus*) are able to track marine temperature gradients with precision^{12,13}.

Compared with mammalian cold receptors, the ampullae of Lorenzini are much more sensitive to abrupt cooling, which may reflect different mechanisms of temperature transduction. Some cold receptors in mice rely on a collaboration of multiple temperature-sensitive ion channels². Although this mechanism may operate in neurons associated with the ampullae, it would not account for the sharks' acute temperature sensitivity. As the gel's ion content is not unique⁷, extracellular thermopower may also be important in other biological systems.

Brandon R. Brown

Department of Physics, University of San Francisco,
San Francisco, California 94117, USA
e-mail: brownb@usfca.edu

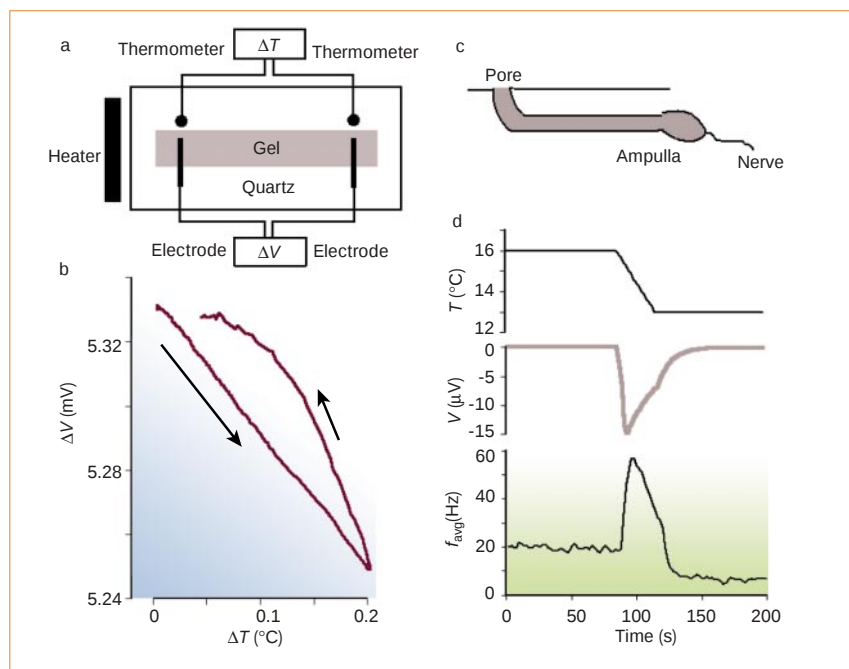


Figure 1 Thermopower and thermal sensitivity of gel samples from shark electro-sensors. **a**, Apparatus used for measuring the gel's thermopower. **b**, The change in voltage with temperature measured in gel extracted from the ampullae of Lorenzini of a black-tip reef shark (samples were collected as described⁷); different curves are obtained for fast heating and subsequent slow cooling (arrows). **c**, Organization of a gel-filled ampulla and canal in the dogfish. **d**, Temperature stimulus and nerve-firing response in dogfish ampulla are shown in the top and bottom curves, respectively (reprinted with permission)⁶; middle curve shows the approximate voltage in the ampulla, calculated from the thermopower data, assuming an exponential decay of the temperature gradient.

1. Reid, G. & Flonta, M.-L. *Nature* **413**, 480 (2001).
2. Viana, F., de la Pena, E. & Belmonte, C. *Nature Neurosci.* **5**, 254–260 (2002).
3. Murray, R. W. *Nature* **187**, 957 (1960).
4. Wilkens, L. A., Russell, D. F., Pei, X. & Gurgens, C. *Proc. R. Soc. Lond. B* **264**, 1723–1729 (1997).
5. Sand, A. *Proc. R. Soc. Lond. B* **125**, 524–553 (1938).
6. Wissing, H., Braun, H. A. & Schafer, K. *Progr. Brain Res.* **74**, 99–110 (1988).
7. Brown, B., Hutchison, J. C., Hughes, M. E., Kellogg, D. R. & Murray, R. W. *Phys. Rev. E* **65**, 061903 (2002).
8. MacDonald, D. K. *Thermoelectricity* (Wiley, New York, 1962).
9. Grosso, G. & Parravicini, G. P. *Solid State Physics* (Academic, London, 2000).
10. Montgomery, J. C. *Comp. Biochem. Physiol.* **79** A, 189–195 (1984).
11. Kalmijn, A. *Nature* **212**, 1232–1233 (1966).
12. Sims, D. W. & Quayle, V. A. *Nature* **393**, 460–464 (1998).
13. Carey, F. G. *et al. Copeia* **2**, 254–260 (1982).

Competing financial interests: declared none.

Nanoelectromechanical systems

Nanodevice motion at microwave frequencies

It has been almost forgotten that the first computers envisaged by Charles Babbage in the early 1800s were mechanical^{1,2} and not electronic, but the development of high-frequency nanoelectromechanical systems is now promising a range of new applications³, including sensitive mechanical charge detectors⁴ and mechanical devices for high-frequency signal processing⁵, biological imaging⁶ and quantum measurement^{7–9}. Here we describe the construction of nanodevices that will operate with fundamental frequencies in the previously inaccessible microwave range (greater than 1 gigahertz). This achievement represents a significant advance in the quest for extremely high-frequency nanoelectromechanical systems.

Until now, it has not been possible to create mechanical devices that operate at extremely high frequencies, owing to the dual challenge of detecting tiny displacements (on the scale of femtometres) at microwave frequencies^{1,3}. The characteristic frequency of nanoelectromechanical systems (NEMS) scales upwards with decreasing size, but their displacement (when operating linearly) and their electro-mechanical impedance both simultaneously scale downwards.

Two advances have been crucial to breaking the 1-GHz barrier in NEMS: the use of silicon carbide epilayers¹⁰, which are of comparable density but are significantly stiffer than the usual silicon^{11,12}, and which allow higher frequencies to be attained for structures of similar geometry; and the development of balanced, high-frequency displacement transducers, which enable the ubiquitous passive embedding impedances that arise from electrical connections to the macroworld to be nulled¹³ (if uncontrolled, these parasitic impedances overwhelm the electromechanical impedance of interest — the 'signal' — in ultrasmall NEMS).

We used 3C–SiC films that were grown hetero-epitaxially at atmospheric pressure by chemical-vapour deposition in an induction-heated reactor on 100-mm-diameter (100) Si wafers¹⁰. Device nanofabrication involves both optical and electron-beam lithography to define, respectively, large-area contact pads and submicrometre-scale, thin metallic-film masks with the device geometry. Pattern transfer to the 3C–SiC layer is achieved by an electron cyclotron resonance (ECR) plasma-etch step involving an NF₃/O₂/Ar mixture. The patterned 3C–SiC beams are then suspended above the underlying silicon substrate by using an isotropic NF₃/Ar ECR etch. The metallic mask

(30 nm of aluminium, followed by 5 nm of titanium), deposited by e-beam evaporation and patterned by lift-off, remains on the beams and is used as the electrode for displacement transduction. The devices consist of two nominally identical, doubly clamped beams, roughly 1.1 µm long, 120 nm wide and 75 nm thick.

Each doubly clamped beam pair is positioned perpendicular to a strong magnetic field (3–8 tesla) *in vacuo* within a liquid-helium cryostat. Balanced magnetomotive detection is used¹³, when the driving frequency matches the fundamental frequency of the in-plane flexural mode for one of the beams, there is resonant enhancement of the induced electromotive force. This response is pre-amplified and characterized by a microwave-network analyser.

Fundamental mechanical resonances are detected at 1.014 GHz and 1.029 GHz for the two beams (Fig. 1). So far, quality factors attained above 1 GHz (about 500) are substantially lower than observed for NEMS in roughly the 100-MHz range (about 10⁴). Having ruled out factors such as electrical damping, we are investigating whether these stems from roughness in the initial SiC epilayers, and how such sources of acoustic loss in microwave NEMS can be minimized. Nonetheless, this step into the previously inaccessible domain of microwave-frequency mechanical excitations constitutes a milestone along the path to the many new applications offered by nanomechanical systems.

Xue Ming Henry Huang*,
Christian A. Zorman†, **Mehran Mehregany†**, **Michael L. Roukes***

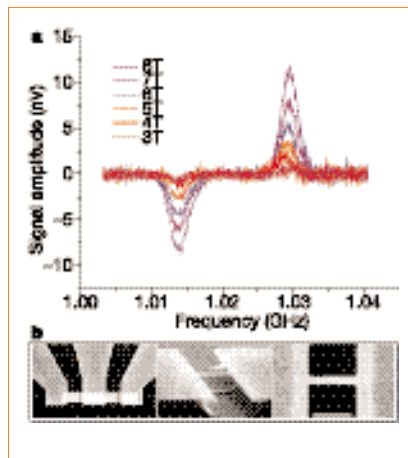


Figure 1 Microwave-frequency nanomechanical devices. **a**, Fundamental flexural-mode resonant mechanical response at 1.014 and 1.029 GHz, detected at about 4.2 K from a pair of doubly clamped silicon carbide beams as a function of applied magnetic field (3–8 tesla). These devices are electrically connected within a balanced magnetomotive detection scheme¹²; each distinct resonance corresponds to excitation of one of the beams within the device. **b**, Scanning electron micrographs of a similar (slightly larger) pair of devices, with magnified views of a single resonant element. Scale bar, bottom right, 1 µm.

*Condensed Matter Physics, California Institute of Technology 114–36, Pasadena, California 91125, USA

e-mail: roukes@caltech.edu

†Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, Ohio 44106, USA

1. Roukes, M. L. *Sci. Am.* **285**, 48–57 (2001).
2. Swade, D. *The Difference Engine: Charles Babbage and the Quest to Build the First Computer* (Viking-Penguin, New York, 2001).
3. Roukes, M. L. *Phys. World* **14**, 25–31 (2001).
4. Cleland, A. N. & Roukes, M. L. *Nature* **392**, 160–162 (1998).
5. Nguyen, C. T.-C., Katehi, L. P. B. & Rebeiz, G. M. *Proc. IEEE* **86**, 1756–1768 (1998).
6. Sidles, J. A. *et al. Rev. Mod. Phys.* **67**, 249–265 (1995).
7. Cho, A. *Science* **299**, 36–37 (2003).
8. Armour, A. D., Blencowe, M. P. & Schwab, K. C. *Phys. Rev. Lett.* **88**, 148301 (2002).
9. Bocko, M. F. & Onofrio, R. *Rev. Mod. Phys.* **68**, 755–799 (1996).
10. Yang, Y. T. *et al. Appl. Phys. Lett.* **78**, 162–164 (2001).
11. Carr, D. W., Evoy, S., Sekarik, L., Craighead, H. G. & Parpia, J. M. *Appl. Phys. Lett.* **75**, 920–922 (1999).
12. Clark, J. R., Hsu, W.-T. & Nguyen, C. T.-C. in *Technical Digest, IEEE Int. Electron Devices Meeting, San Francisco, California (11–13 December 2000)* 493–496 (2001).
13. Ekinci, K. L., Yang, Y. T., Huang, X. M. H. & Roukes, M. L. *Appl. Phys. Lett.* **81**, 2253–2255 (2002).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Genomic function

Rate of evolution and gene dispensability

Whether more dispensable genes evolve faster than less dispensable ones¹ is a contentious issue^{2–4}. Comparing yeast and worm genes, Hirsh and Fraser³ observe a gradual tendency for less dispensable genes (those that reduce the growth rate of yeast when knocked out) to have lower rates of protein evolution. Here we repeat their analysis using larger data sets and find no evidence that dispensability explains the variation in rates of protein evolution. Although Hirsh and Fraser provide a model to show why their result is to be expected, our analysis suggests that their model, which assumes among other things that no substitution is advantageous, cannot be generally applied.

To estimate the rate of evolution, Hirsh and Fraser used orthologues identified in the worm *Caenorhabditis elegans*. But using so distant a relative of the yeast *Saccharomyces cerevisiae* may have pitfalls. We have therefore repeated the analysis using *C. elegans* and three other closer relatives of *S. cerevisiae* (*S. bayanus*, *Candida albicans* and *Schizosaccharomyces pombe*). Unlike *Caenorhabditis*, for which the most recent common ancestor with yeast existed between 1.0 billion and 1.6 billion years ago, these three species are separated from *S. cerevisiae* by about 10 million, 140–310 million and 330–440 million years, respectively. New fitness data⁵ enable complete genomes to be analysed using samples that

are, on average, an order of magnitude larger than those used by Hirsh and Fraser.

Orthologues of *S. cerevisiae* proteins in each of the four genomes were identified by reciprocal pairwise searches⁶. Protein sequences were aligned using CLUSTAL-W software with the default settings; only those alignments with less than 20% gap were retained. Protein evolutionary distances were calculated by using a maximum-likelihood method⁷ and Dayhoff's substitution matrix (using different substitution matrices or allowing for rate variation among sites did not alter our results). Whole-genome transcription data⁸ gave a measure of gene expression (using codon-adaptation indexes as estimates of gene-expression rates⁹ gives results that are qualitatively similar).

The correlation between dispensability (that is, the effect on fitness of knocking out the gene) and the rate of evolution for each of the analyses suggests, at most, only a very weak effect (Fig. 1). Although, in Hirsh and Fraser's original analysis, gene dispensability accounts for 8.5% (or 20% under ranked correlation) of the variation in rate of protein evolution (A. Hirsh, personal communication), in the larger data sets this is reduced by roughly an order of magnitude (Fig. 1); Hirsh and Fraser's repeat worm analysis showed a similar reduction¹⁰. This indicates that the high r^2 value obtained in the authors' original analysis is largely an artefact due to their limited sample size.

Likewise, the variation in the correlation coefficients in our four data sets is a result of the different sets of appropriate orthologues used. When the same set of *S. cerevisiae* genes is used for comparison with

C. albicans, *S. pombe* and *C. elegans*, the correlations are not statistically different and none is significant (results not shown).

The tiny amount of variation that might be explained by dispensability seems to be a weak covariate of expression rate. More dispensable genes tend to be expressed at a lower rate than less dispensable ones ($N=3,783$, Pearson $r_{\log(\text{expression})-\text{fitness}}=0.191$, $P<10^{-8}$, Spearman $\rho_{\text{expression}-\text{fitness}}=0.155$, $P<10^{-8}$), possibly because these genes have less phenotypic contribution. The rate of expression is a good predictor of the rate of protein evolution¹¹. If we control for the covariance with expression rates, we do not see any correlation between fitness and rate of protein evolution ($P>0.172$ for all comparisons; Fig. 1 legend). However, when we control for the dispensability of genes, we still find that expression rate is a strong predictor (Spearman partial $\rho_{\text{expression}-\text{protein distance}|\text{fitness}}<-0.49$, $P<10^{-7}$ for all comparisons).

We therefore see no evidence that dispensability is a relevant variable in understanding the rates of protein evolution. More generally, claims that any given variable is important in predicting rates of protein evolution should control for what seems to be a good predictor — the rate of expression. Neither gene dispensability nor recombination rate⁶ seems to be as important a variable after such control.

Csaba Pál*, **Balázs Papp***,
Laurence D. Hurst*

*Department of Biology and Biochemistry,
University of Bath, Bath BA2 7AY, UK
e-mail: l.d.hurst@bath.ac.uk

†Department of Plant Taxonomy and Ecology,
Eötvös University, Pázmány Péter Sétány 1/C,
1117 Budapest, Hungary

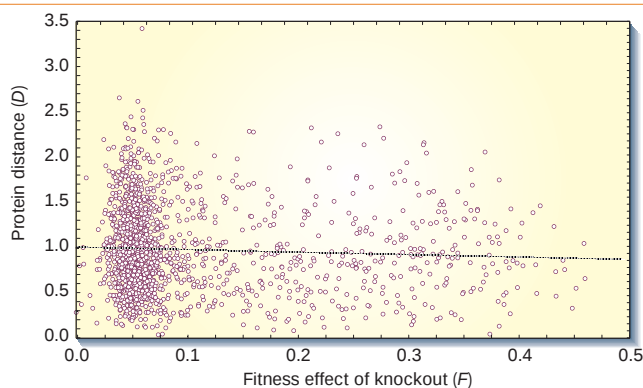


Figure 1 Relationship between the rate of protein evolution (protein distance, D) and the fitness effect of knockout (F , 1 – dispensability) for genes in a comparison involving a range of yeast species ($N=1,660$, Pearson $r^2_{FD}=0.00289$, $P=0.028$, Spearman rank $\rho^2_{FD}=0.00515$, $P=0.0034$). Graph shows the result of the comparison of *Saccharomyces cerevisiae* and *Candida albicans*; dotted line indicates the best-fit linear regression. After controlling for expression rate (E), the correlation disappears (for *C. albicans*: $N=1,554$, Pearson partial $r_{FD|E}=0.003$, $P=0.9$, Spearman partial rank $\rho_{FD|E}=-0.0343$, $P=0.172$; for *Schizosaccharomyces pombe*: $N=1,090$, $r_{FD}=-0.103$, $P<0.001$, $r_{FD|E}=-0.015$, $P=0.619$, $\rho_{FD}=-0.1077$, $P<0.001$, $\rho_{FD|E}=-0.0249$, $P=0.41$; for *Caenorhabditis elegans*: $N=490$, $r_{FD}=-0.122$, $P=0.007$, $r_{FD|E}=0.0134$, $P=0.768$, $\rho_{FD}=-0.171$, $P<0.001$, $\rho_{FD|E}=-0.0449$, $P=0.317$; for *Saccharomyces bayanus*: $N=154$, $r_{FD}=-0.151$, $P=0.061$, $r_{FD|E}=-0.0159$, $P=0.845$, $\rho_{FD}=-0.227$, $P=0.0047$, $\rho_{FD|E}=-0.0864$, $P=0.294$). P values for Spearman partial-rank correlations were assessed by randomization. Only non-essential genes were studied ($F<1$), as in ref. 3.

1. Wilson, A. C., Carlson, S. S. & White, T. J. *Annu. Rev. Biochem.* **46**, 573–639 (1977).
2. Hurst, L. D. & Smith, N. G. C. *Curr. Biol.* **9**, 747–750 (1999).
3. Hirsh, A. E. & Fraser, H. B. *Nature* **411**, 1046–1049 (2001).
4. Jordan, I. K., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. *Genome Res.* **12**, 962–968 (2002).
5. Steinmetz, L. M. *et al.* *Nature Genet.* **31**, 400–404 (2002).
6. Pál, C., Papp, B. & Hurst, L. D. *Mol. Biol. Evol.* **18**, 2323–2326 (2001).
7. Yang, Z. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
8. Holstege, F. C. *et al.* *Cell* **95**, 717–728 (1998).
9. Coghlan, A. & Wolfe, K. H. *Yeast* **16**, 1131–1145 (2000).
10. Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C. & Feldman, M. W. *Science* **296**, 750–752 (2002).
11. Pál, C., Papp, B. & Hurst, L. D. *Genetics* **158**, 927–931 (2001).
12. Gibbons, J. D. *Nonparametric Measures of Association* (ed. Lewis-Beck, M. S.) (Sage, Newbury Park, UK, 1993).

Hirsh and Fraser reply — The relationship between protein dispensability and rate of evolution that we detected in yeast¹ has since been confirmed among bacteria² — as well as in the same data set^{3,4} that Pal *et al.* refer to as “new fitness data” and which they re-analyse here using methods that fail to reveal the relationship.

In each of the three studies that demonstrate this association, the correlation was small but highly significant. The interest of the finding was not that it explained a large proportion of the total variance in protein evolutionary rate (it did not), but rather that, despite several obvious sources of noise and error — for example, fitness effects were measured in a single laboratory medium and not in the wild — a statistically significant pattern was detectable, confirming a basic prediction of a widely held model of protein evolution⁵. The question that arises from the comment of Pal *et al.* is therefore whether there is still a significant relationship between dispensability and evolutionary rate when the best available estimates of evolutionary rate, expression level and deletion effect are used.

In a re-analysis that differs from that of Pal *et al.* in several important ways, we still find a highly significant relationship between protein dispensability and evolutionary rate, even when we control for the level of gene expression. Using three complete genomes in the *Saccharomyces* genus, each of which diverged from *S. cerevisiae* less than 10 million years ago, as well as the more distant comparisons used by Pal *et al.* (*S. pombe*, *C. albicans* and *C. elegans*), we estimated the evolutionary rates of *S. cerevisiae* proteins. Pal *et al.* relied exclusively on BLAST to identify putative orthologues. However, the closest BLAST hit is often not the nearest evolutionary neighbour⁶, and this problem is exacerbated when using an incomplete genome, as Pal *et al.* did. We therefore first estimated maximum-likelihood evolutionary distances⁷ for multiple, highly significant BLAST hits in the complete genomes, and then used these distances to identify putative orthologues (algorithm by courtesy of D. P. Wall).

Pal *et al.* used a data set of 4,273 deletion effects, each of which is based on two replicate measurements; we used a new data set of 5,937 deletion effects, 4,744 of which are based on more than 10 replicate measurements (data by courtesy of J. Kumm and G. Giaever). Calculating Kendall's correlation coefficient⁸ between deletion effect and evolutionary distance, we observed a highly significant negative relationship ($-0.26 < \tau < -0.17$, $P < 0.0001$ in all evolutionary comparisons; manuscript in preparation). To control for different levels of gene expression, we used recent expression data⁹ that were not measured in an aneuploid strain of yeast¹⁰. Calculating Kendall's partial correlation coefficient⁸ ($-0.18 < \tau < -0.16$, $P < 0.0001$ in all evolutionary comparisons), we find that the relationship between protein dispensability and evolutionary rate remains highly significant, even when controlling for gene-expression levels.

Aaron E. Hirsh*, Hunter B. Fraser†

*Department of Biological Sciences, Stanford University, Stanford, California 94305, USA
e-mail: aehirsh@stanford.edu

†Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

- Hirsh, A. E. & Fraser, H. B. *Nature* **411**, 1046–1049 (2001).
- Jordan, I. K., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. *Genome Res.* **12**, 962–968 (2002).
- Steinmetz, L. M. *et al.* *Nature Genet.* **31**, 400–404 (2002).
- Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C. & Feldman, M. W. *Science* **296**, 750–752 (2002).
- Wilson, A. C., Carlson, S. S. & White, T. J. *Annu. Rev. Biochem.* **46**, 573–639 (1977).
- Koski, L. B. & Golding, G. B. *J. Mol. Evol.* **52**, 540–542 (2001).
- Yang, Z. *Phylogenetic Analysis by Maximum Likelihood Version 3.12* (Univ. Coll. Lond., London, 2000).
- Gibbons, J. D. (ed.) *Nonparametric Measures of Association* (Sage, Newbury Park, UK, 1993).
- Wang, Y. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 5860–5865 (2002).
- Hughes, T. R. *et al.* *Nature Genet.* **25**, 333–337 (2000).

COMMUNICATIONS ARISING

Cosmology

Do black holes constrain varying constants?

Tentative observations of shifted spectral lines in distant quasars^{1,2} have rekindled interest in Dirac's old idea^{3,4} that the fundamental 'constants' of physics may vary over time. Davies *et al.* have argued that black-hole thermodynamics favours theories in which the speed of light, c , decreases, and does not favour those in which the fundamental electronic charge, e , increases⁵. Here we show, however, that when the entire thermal environment of a black hole is considered, no such conclusion can be drawn. Although black-hole features such as mass quantization may still constrain models with varying 'constants'⁶, thermodynamics probably cannot.

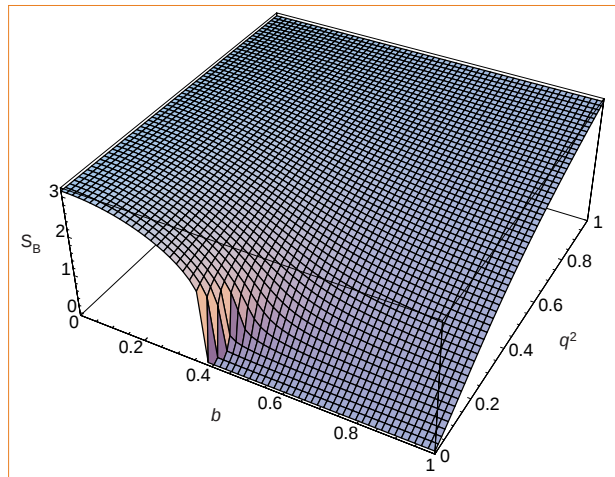


Figure 1 Entropy, S_B , of a black hole and its surroundings as a function of charge, q^2 , and inverse temperature, b . For any fixed temperature, an increase in the fine-structure constant increases q^2 and therefore S_B .

The observations of refs 1, 2 suggest that the fine-structure constant, α , may have been slightly smaller in the early Universe. As $\alpha = e^2/\hbar c$ depends on e , c and Planck's constant, $\hbar$, it is natural to ask which of these values varies. Davies *et al.* offer an ingenious argument. A black hole with mass M and charge $Q = ne$ has an entropy⁷ $S = \pi G/\hbar c [M + (M^2 - n^2 e^2/G)^{1/2}]^2$. Evidently, an increase in e will reduce the entropy, violating the generalized second law of thermodynamics, whereas a decrease in $\hbar$ or c will increase the entropy. This argument involves assumptions that may not be valid for all models^{6,8,9}, but it offers an interesting starting point.

As Davies *et al.* note, however, such an argument should consider not just the black hole, but also its surroundings. An isolated black hole is never in thermal equilibrium: it decays by Hawking radiation and, if it is charged, by spontaneous emission of charged particles¹⁰. These processes reduce S , but do not violate the second law of thermodynamics because there is a compensating increase in the entropy of the environment.

To investigate the thermodynamics of varying 'constants', one should study a black hole that is in equilibrium with its environment. This can be done by considering a black hole in a 'box' of radius r_B , with fixed boundary temperature T and charge Q (the canonical ensemble) or electrostatic potential ϕ (the grand canonical ensemble). Note that r_B can be altered only by doing work on the system.

In the canonical ensemble, the entropy is given by $S = \pi r_B^2 x^2$, where x is determined by the seventh-order equation¹¹ $x^5(x - q^2)(x - 1) + b^2(x^2 - q^2)^2 = 0$, where $q = \sqrt{GQ/r_B c^2}$ and $b = \hbar c/4\pi r_B kT$. Figure 1 shows a plot of S/r_B^2 against q^2 and b . It is apparent — and may be confirmed numerically — that the entropy increases with increasing α . For the grand canonical ensemble, exact analytical results lead to the same conclusion. Black-hole thermo-

dynamics thus militates against models in which the fundamental charge, e , decreases, but places no restriction on increasing e .

To compare this result with that of Davies *et al.*, note first that the Hawking temperature of a charged black hole decreases with increasing e . A black hole will thus cool below the ambient temperature of the heat bath and will absorb heat, thereby increasing its mass. According to the first law of thermodynamics, the net change in entropy is $dS = 1/T(dE - \phi dQ)$, and it may be verified that the increase in the energy, E , dominates.

Of course, such thermodynamic arguments only describe relationships among equilibria, and not the transitions between equilibria. A more detailed analysis, however, requires an explicit, dynamic model. In particular, any theory with a variable fine-structure constant necessarily contains a new scalar field, α itself, the entropy of which cannot be neglected during dynamic processes in which α is changing. Jacobson (personal communication) has suggested that a suitable dynamic version of the second law of thermodynamics¹² will ensure that entropy increases during such a process. Black-hole thermodynamics is therefore insufficient to constrain theories in which α increases.

Steven Carlip, Sachindeo Vaidya

Department of Physics, University of California, Davis, California 95616, USA
e-mail: carlip@dirac.ucdavis.edu

- Webb, J. K. *et al.* *Phys. Rev. Lett.* **82**, 884–887 (1999).
- Webb, J. K. *et al.* *Phys. Rev. Lett.* **87**, 091301 (2001).
- Dirac, P. A. M. *Nature* **139**, 323 (1937).
- Dirac, P. A. M. *Proc. R. Soc. Lond. A* **165**, 199–208 (1938).
- Davies, P. C. W., Davis, T. M. & Lineweaver, C. H. *Nature* **418**, 602–603 (2002).
- Carlip, S. *Phys. Rev. D* **67**, 023507 (2003).
- Hawking, S. W. *Commun. Math. Phys.* **43**, 199–220 (1975).
- Flambaum, V. V. eprint astro-ph/0208384.
- Barrow, J. D. eprint gr-qc/0211074.
- Gibbons, G. W. *Commun. Math. Phys.* **44**, 245–264 (1975).
- Braden, H. W., Brown, J. D., Whiting, B. F. & York, J. W. *Phys. Rev. D* **42**, 3376–3385 (1990).
- Jacobson, T., Kang, G. & Myers, R. C. *Phys. Rev. D* **52**, 3518–3528 (1995).

DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation

Christopher J. Bakkenist & Michael B. Kastan

Department of Hematology–Oncology, St Jude Children's Research Hospital, 332 North Lauderdale Street, Memphis, Tennessee 38105, USA

The ATM protein kinase, mutations of which are associated with the human disease ataxia–telangiectasia, mediates responses to ionizing radiation in mammalian cells. Here we show that ATM is held inactive in unirradiated cells as a dimer or higher-order multimer, with the kinase domain bound to a region surrounding serine 1981 that is contained within the previously described 'FAT' domain. Cellular irradiation induces rapid intermolecular autophosphorylation of serine 1981 that causes dimer dissociation and initiates cellular ATM kinase activity. Most ATM molecules in the cell are rapidly phosphorylated on this site after doses of radiation as low as 0.5 Gy, and binding of a phosphospecific antibody is detectable after the introduction of only a few DNA double-strand breaks in the cell. Activation of the ATM kinase seems to be an initiating event in cellular responses to irradiation, and our data indicate that ATM activation is not dependent on direct binding to DNA strand breaks, but may result from changes in the structure of chromatin.

Eukaryotic cells have evolved complex mechanisms to deal with environmental stresses. Signal-transduction pathways are rapidly activated after exposure to DNA-damaging agents and other cellular stresses, and these pathways affect processes such as gene transcription and cell-cycle progression^{1–3}. ATM, the gene that is mutated in the human disease ataxia–telangiectasia (AT), is crucial for the initiation of signalling pathways in mammalian cells following exposure to ionizing radiation (IR) and to other agents that introduce double-strand breaks into cellular DNA^{4,5}. Cells from AT patients typically lack detectable ATM protein, contain abnormalities in telomere morphology and have abnormal responses to IR, including increased cell death, increased chromosomal breakage and cell-cycle checkpoint defects⁶. In addition, AT patients exhibit progressive cerebellar ataxia, immune deficiencies, gonadal atrophy, oculocutaneous telangiectasias, radiation sensitivity, premature ageing and increased risk of cancers, particularly lymphomas.

The ATM gene encodes a 370-kDa protein that belongs to the phosphoinositide 3-kinase (PI(3)K) superfamily⁷, but which phosphorylates proteins rather than lipids^{8,9}. The 350-amino-acid kinase domain at the carboxy terminus of this large protein is the only segment of ATM with an assigned function. Exposure of cells to IR triggers ATM kinase activity, and this function is required for arrests in G1, S and G2 phases of the cell cycle⁵. Several substrates of the ATM kinase participate in these IR-induced cell-cycle arrests. These include p53, Mdm2 and Chk2 in the G1 checkpoint^{8–12}; Nbs1, Brca1, FancD2 and SMC1 in the transient IR-induced S-phase arrest^{13–19}; and Brca1 and hRad17 in the G2/M checkpoint^{20,21}.

The mechanisms by which eukaryotic cells sense DNA strand breaks remain to be elucidated, but the rapid induction of ATM kinase activity following IR suggests that it acts at an early stage of signal transduction in mammalian cells^{8,9}. Transfected ATM is a phosphoprotein, raising the possibility that ATM kinase activity is modulated by post-translational modification. We have now identified an important site of ATM phosphorylation induced by IR; we have shown that it results from intermolecular autophosphorylation, and have proceeded to explore the functional significance of the phosphorylation event. This phosphorylation event does not directly regulate the activity of the kinase, but instead disrupts ATM oligomers, which in turn allows accessibility of substrates to the

ATM kinase domain. The rapidity and stoichiometry of the reaction indicate that ATM is not activated by binding directly to DNA strand breaks, and we present data that support a model in which DNA damage rapidly causes changes in higher-order chromatin structures that initiate ATM activation.

Ionizing radiation induces autophosphorylation of ATM

Exposure of cells to IR significantly increased the incorporation of ³²P-orthophosphate into both transfected ATM (Fig. 1a) and endogenous ATM (Fig. 1b). No such increase was observed after the irradiation of cells that had been transfected with kinase-inactive ATM. Incorporation of radioactive phosphate into ATM also occurs in *in vitro* assays of ATM kinase activity^{9,22}. This *in vitro* phosphorylation of ATM is not seen with the kinase-inactive ATM protein, is inhibited by exposure to 30 nM wortmannin, is dependent on the addition of manganese, and is not dependent on the addition of exogenous DNA (Fig. 1c). Identical properties are characteristic of phosphorylation of target substrates by ATM^{9,22,23}. These observations are all consistent with a model in which ATM phosphorylates itself. To identify the site(s) of phosphorylation in ATM, radioactively labelled ATM protein was digested with trypsin, and tryptic phosphopeptides were evaluated by sequential two-dimensional electrophoresis and chromatography. A single major *de novo* phosphorylated peptide was identified in both transfected (Fig. 1d) and endogenous (Fig. 1e) ATM isolated from irradiated cells. Phospho-amino-acid analysis showed the target amino acid to be a serine (data not shown).

After failing to identify the target peptide by automated sequencing, mass spectrometry, predicted chromatographic migration, and mutagenesis of all of the serines in the ATM kinase domain, the labelled phosphopeptide was ultimately identified by secondary proteolytic and chemical cleavage of the primary phosphopeptide. Digestion of the primary tryptic phosphopeptide with V8 protease resulted in a peptide with altered chromatographic mobility (Fig. 1f), dictating the presence of a glutamic acid in the peptide. Unchanged peptide mobility after cyanogen bromide treatment in formic acid indicated that it does not contain methionine (data not shown). Manual Edman degradation of both the principal tryptic phosphopeptide and the V8/tryptic phosphopeptide

revealed a secondary spot in each cycle, consistent with a derivatized C-terminal lysine ϵ -amino group peptide²⁴ (Fig. 1g, h). The phosphorylated serine was released after the eighth or ninth cycle of Edman degradation from the tryptic peptide and after the second cycle from the V8/tryptic peptide (Fig. 1g, h). Thus, the phosphorylated serine is eight or nine amino acids to the C-terminal side of either a lysine or arginine residue in the tryptic peptide, and two amino acids to the C-terminal side of a glutamic acid residue in the V8/tryptic peptide. The only tryptic peptide in ATM that meets these sequence requirements is the 19-residue peptide 1974-SLA-FEEGSPQSTTISLSEK-1992, with serine 1981 as the predicted phosphorylated serine (Fig. 1i). Serine 1981 is conserved in mouse and *Xenopus* ATM (Fig. 1i), but is not found in suspected homologues of less complex metazoans, and is located in the amino terminus of the FAT domain, a region of approximately 500 amino acids with some conservation across the PI(3)K family of kinases, including Frap, ATM and Trapp²⁵. Because the target serine is in an 'SQ' site, either an autophosphorylation event or phosphorylation by an ATM family member is indicated²².

Rabbit polyclonal antibodies were generated that specifically recognize serine 1981 only when it is unphosphorylated (anti-1981S) or only when it is phosphorylated (anti-1981S-P), and specificity was shown on peptides and full-length protein (Fig. 2a–c). Within 30 min after the irradiation of cells at 10 Gy, the recognition of wild-type ATM by the anti-1981S antisera decreased

several fold, whereas recognition of kinase-inactive ATM was unchanged (Fig. 2c). Conversely, binding of the anti-1981S-P antisera increased several fold after IR, whereas recognition of kinase-inactive ATM was unaffected. The decrease in anti-1981S binding indicates that a high fraction of total cellular ATM becomes modified on serine 1981 after IR, and that anti-1981S-P binding results mirror the metabolic labelling of transfected wild-type ATM and kinase-inactive ATM in 293T cells (Fig. 1a). Phosphorylation of serine 1981 in endogenous ATM in non-transformed human fibroblasts was not apparent in unperturbed cells, but was readily detectable 1 h after exposure to IR and 5 h after exposure to both IR and ultraviolet irradiation (Fig. 2d). Recognition of ATM by the phosphospecific antibody after IR cannot simply be due to changes in cell-cycle distribution, because cell-cycle profiles do not change significantly in the first hour after IR, and this phosphorylation event is also seen in primary fibroblasts that are arrested in G0 after either IR or ultraviolet irradiation (Fig. 2c, right column).

ATM autophosphorylation had been suggested by the dependence of the enhanced incorporation of phosphate after IR on ATM kinase activity *in vivo* and *in vitro*, by its wortmannin sensitivity, and by the fact that the target serine was an 'SQ' site. A polypeptide containing the appropriate serine residue at 1981 was an excellent *in vitro* substrate of the ATM kinase (Fig. 2e), showing that ATM could phosphorylate this site. In addition, wortmannin concentrations of 20 μ M or more effectively inhibited phosphorylation of

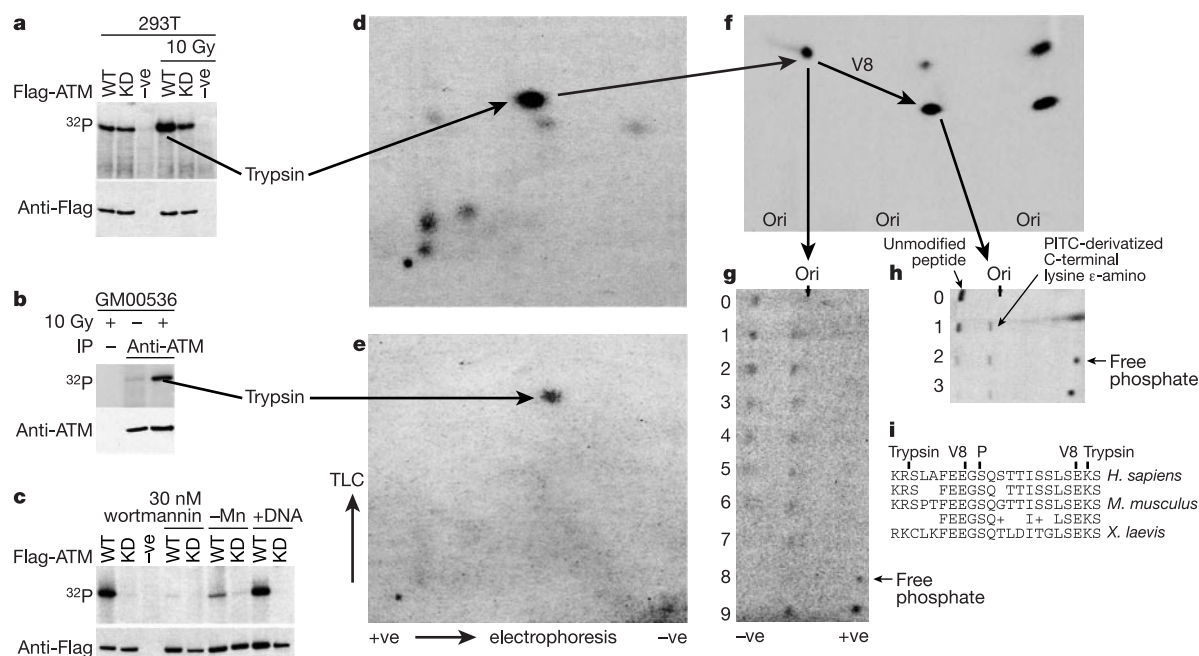


Figure 1 Metabolic labelling of ATM in response to ionizing radiation. **a**, ATM phosphorylation is increased in response to ionizing radiation (IR) and is dependent on ATM kinase activity. After transfection of Flag-tagged wild-type (WT) or mutant (KD) ATM, 293T cells were irradiated (10 Gy) and then incubated with ³²P-orthophosphate for 30 min. Radiolabelled ATM was immunoprecipitated with anti-Flag, and evaluated by autoradiography (upper panel) and by immunoblot with anti-Flag antibody (lower panel). **b**, Phosphorylation of endogenous ATM is increased in response to IR. GM00536 cells were irradiated (10 Gy) and then metabolically labelled with ³²P-orthophosphate for 30 min. ATM was immunoprecipitated (IP) and evaluated by autoradiography (upper panel) and by immunoblot with anti-ATM antibody (lower panel). **c**, WT-ATM autophosphorylates *in vitro*. After transfection into 293T cells, Flag-tagged WT-ATM or KD-ATM were immunoprecipitated with anti-Flag and used in *in vitro* kinase assays under different conditions. Wortmannin was added to the samples indicated, on ice, 15 min before the kinase reaction, and 5 μ M manganese and 1 μ g sonicated calf thymus DNA were present or absent in the kinase reaction, as indicated. **d**, **e**, A single principal tryptic

phosphopeptide in ATM is detected after IR. Flag-tagged WT-ATM or endogenous ATM labelled following IR was digested on the nitrocellulose membrane with trypsin. Peptides were purified and spotted at the origin, and subjected to electrophoresis (x axis) and thin-layer chromatography (TLC; y axis). **f**, The principal tryptic phosphopeptide contains a glutamic acid residue. Peptides generated by V8 protease digestion of the principal tryptic ATM phosphopeptide were spotted (Ori) on the plate, left to right, as follows: purified tryptic peptide, V8 digested tryptic peptide and a mixture of the two samples. **g**, **h**, Edman degradation. The principal tryptic peptide and the V8/tryptic peptide were purified and subjected to the noted number of cycles of manual Edman degradation. Phenyl isothiocyanate (PITC) derivatization of the lysine ϵ -amino group results in changed mobility of the peptides after one cycle. Free phosphate is released from the tryptic or V8/tryptic phosphopeptides after eight to nine cycles or after two cycles, respectively. **i**, The predicted IR-induced site of phosphorylation, serine 1981, is conserved in mouse and *Xenopus*. –ve, negative; +ve, positive.

serine 1981 after the irradiation of human diploid fibroblasts (Fig. 2f). This concentration of wortmannin inhibits both ATM and DNA-dependent protein kinase (DNA-PK), but not ATR (ataxia-telangiectasia and Rad3 related), *in vivo*²³. The kinetics and levels of IR-induced phosphorylation of serine 1981 were identical in matched cell lines with and without DNA-PK activity (data not shown), thus excluding DNA-PK as a potential responsible kinase.

Although the previous *in vivo* experiments here had shown that optimal phosphorylation of ATM was dependent on the presence of active ATM kinase, their interpretation was complicated by the fact that the cells that we used contained endogenous wild-type ATM. Even the kinase-inactive mutant of ATM can be phosphorylated to some extent in these cells (Figs 1a and 2c). To further demonstrate the importance of ATM kinase activity in the phosphorylation of serine 1981, constructs encoding wild-type, kinase-inactive and S1981A-ATM were transfected into AT cells so that the only potential source of ATM kinase activity was the transgene being used. All of these constructs were expressed at similar levels in AT cells, but only wild-type ATM was recognized by anti-1981S-P, and its binding was increased several fold within 30 min of exposure to IR at 10 Gy (Fig. 2g). Thus, although some phosphorylation of kinase-inactive ATM is observed after transfection into 293T cells

(which contain endogenous ATM activity), transfection into a cell line that lacks ATM kinase activity shows no detectable phosphorylation on serine 1981. Therefore, phosphorylation of serine 1981 depends on the activity of the ATM kinase itself, and the phosphorylation of transfected kinase-inactive ATM must occur *in trans*.

Dissociation of ATM into active monomers

Because the various ATM constructs had been introduced into AT cells, the functional ramifications of serine 1981 phosphorylation could also begin to be assessed. Although the introduction of wild-type ATM restored IR-induced p53 phosphorylation to AT cells, neither kinase-inactive nor S1981A-ATM provided this functional complementation (Fig. 2g). By contrast, the *in vitro* kinase activity of S1981A-ATM was indistinguishable from wild-type ATM (Fig. 3a). Transfection of S1981A-ATM into HeLa cells inhibited both the IR-induced G2 checkpoint (Fig. 3b) and S-phase replication arrest (Fig. 3c) in a manner indistinguishable from the kinase-inactive ATM construct. Consistent with its abrogation of the IR-induced cell-cycle checkpoints, the transfected S1981A mutant also blocked the IR-induced phosphorylation of endogenous ATM protein on serine 1981 (Fig. 3d). Thus, although mutation of serine 1981 did not abolish ATM kinase activity *in vitro*, expression of S1981A-ATM fails to complement AT cells, and effectively inhibits

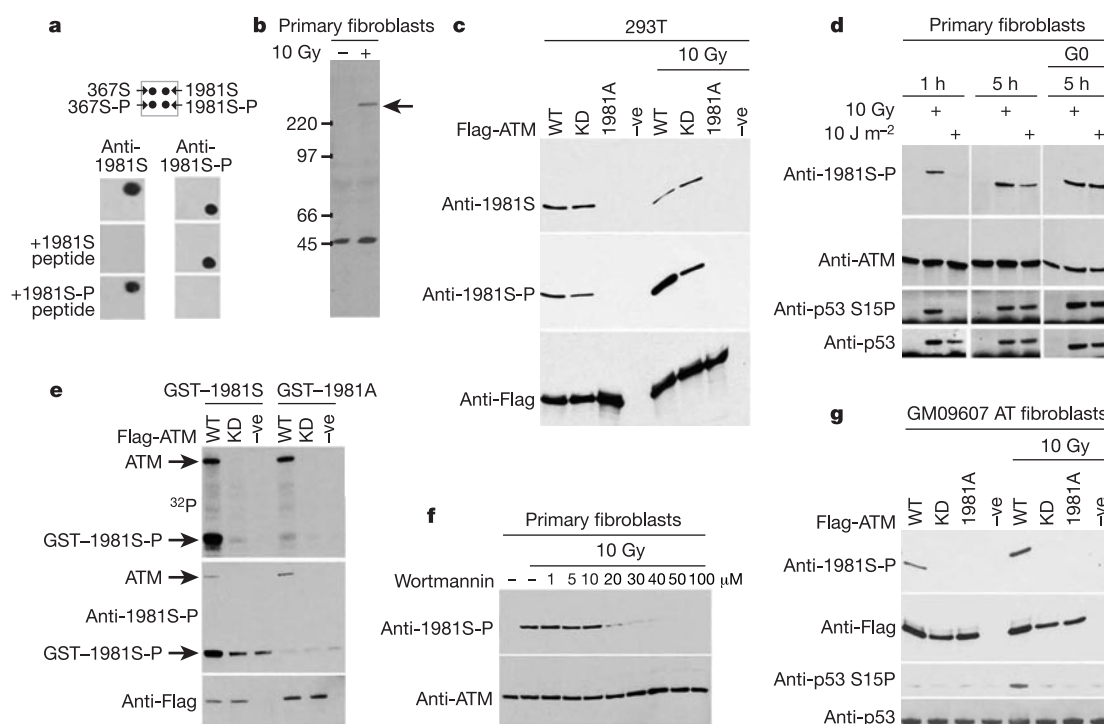


Figure 2 Intermolecular autophosphorylation of ATM serine 1981 occurs *in vivo* after irradiation. **a**, Anti-1981S and anti-1981S-P specific antibodies were generated. Dot blots were prepared with 1 μ g spots of 367S, 1981S, 367S-P and 1981S-P peptides, as indicated in the schema, and were immunoblotted with either anti-1981S or 1981S-P specific antisera. To confirm specificity, 1981S or 1981S-P peptide (20 μ g ml⁻¹) was added as a blocker to the antisera, as indicated, for 1 h before immunoblotting. **b**, Anti-1981S-P recognizes a single protein in crude extracts prepared from exponentially growing primary fibroblasts that is induced by IR (arrow). Whole-cell extract was prepared 30 min after 10 Gy IR, and 10 μ g of each sample was resolved and immunoblotted with anti-1981S-P antibody. **c**, 1981S is phosphorylated *in vivo*. Thirty minutes after irradiation (10 Gy), transfected Flag-tagged WT-ATM, KD-ATM or S1981A-ATM was immunoprecipitated from 293T cells and immunoblotted with anti-1981S, anti-1981S-P or anti-Flag antibodies. **d**, ATM is phosphorylated *de novo* in primary fibroblasts following exposure to IR and ultraviolet light. Exponentially dividing and G0 primary fibroblasts were exposed to either 10 Gy IR or 10 J m⁻² ultraviolet, and whole-cell extracts were prepared

after 1 h and 5 h. ATM was immunoprecipitated and the samples were blotted with anti-1981S-P and anti-ATM antibodies. Levels of p53 serine 15 phosphorylation and total p53 were determined by immunoblot. **e**, ATM phosphorylates a glutathione *S*-transferase (GST) fusion protein containing 1981S *in vitro*. GST fusion proteins containing 1981S or 1981A were generated, purified and used as substrates in *in vitro* kinase assays with either WT or KD Flag-ATM. Upper panel, ³²P incorporation; middle panel, 1981S-P immunoblot; lower panel, Flag-ATM immunoblot. **f**, IR-induced serine 1981 phosphorylation is inhibited by 20 μ M wortmannin. Primary fibroblasts were incubated with wortmannin for 1 h and exposed to 10 Gy IR; serine 1981 phosphorylation was assessed in ATM immunoprecipitated from whole-cell extracts that were prepared 30 min later. **g**, ATM kinase activity is required for serine 1981S phosphorylation after IR. Flag-tagged WT-ATM, KD-ATM or S1981A-ATM was expressed in ataxia-telangiectasia (AT) fibroblasts. Thirty minutes after 10 Gy IR, ATM was immunoprecipitated and immunoblotted with anti-1981S-P and anti-Flag antibodies. Complementation of IR-induced p53 phosphorylation was also assessed (bottom panels).

the cellular activities of endogenous ATM in a dominant-inhibitory manner.

To investigate the reasons for the different effects of serine 1981 mutation on *in vitro* versus cellular activities of ATM, biochemical studies of ATM domains and protein–protein interactions were pursued. Although a peptide containing the ATM kinase domain was insoluble in bacteria when co-transfected with glutathione S-transferase (GST) alone or with a GST fusion with the ATM target peptide p53(1–101), a significant amount of the peptide was stabilized and solubilized in the presence of a GST fusion protein containing ATM amino acids 1961–2046 (Fig. 4a, top panel). The kinase domain was also solubilized in the presence of the GST 1961–2046 fusion protein containing S1981A, but remained insoluble in the presence of phosphorylation-mimic fusion peptides S1981D and S1981E. Furthermore, the soluble kinase domain co-purified on glutathione-agarose with the GST 1961–2046 fusion proteins containing wild-type sequence or S1981A (Fig. 4a, third panel). These results show that the kinase domain and phosphorylation domain can stably bind to one another, and that sequences flanking serine 1981 are crucial for this interaction. The fact that the interaction is prevented by mutation of serine 1981 to either aspartic acid or glutamic acid, both of which have charged side chains that mimic serine phosphorylation, indicated that phosphorylation of serine 1981 would prevent the interaction of this domain with the kinase domain (see below).

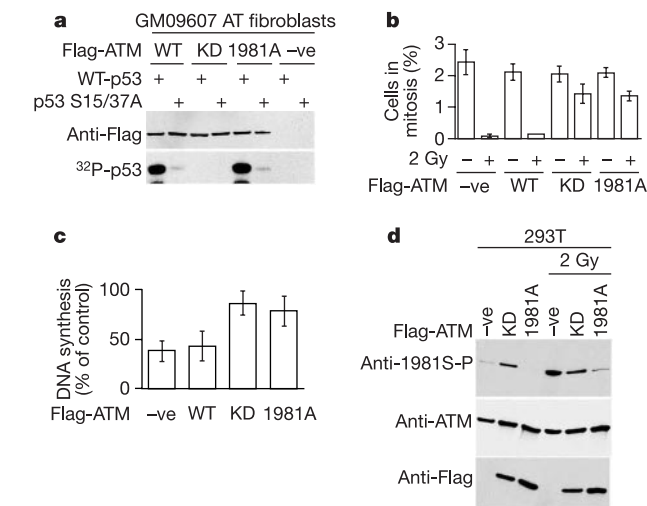


Figure 3 S1981A-ATM has kinase activity but is dominant-inhibitory in cells. **a**, Flag-tagged S1981A-ATM has kinase activity. Flag-tagged WT-ATM, KD-ATM or S1981A-ATM was expressed in AT fibroblasts and immunoprecipitated for use in *in vitro* kinase assays using GST-p53 substrates. 32 P incorporation into both GST-WT-p53 and GST-p53 mutated to alanine at the two 'SQ' sites, serines 15 and 37, was assessed. **b**, Flag-tagged S1981A-ATM is dominant-inhibitory in a G2/M checkpoint assay. The three Flag-tagged ATM constructs or vector alone were transfected into HeLa cells. The percentage of cells in mitosis with and without 2 Gy IR was assessed using flow cytometry by examining histone H3 serine 10 phosphorylation and DNA content. **c**, Flag-tagged S1981A-ATM is dominant-inhibitory in an S-phase checkpoint assay. HeLa cells pre-labelled with 14 C-thymidine were transfected with the three Flag-tagged ATM constructs or vector alone, irradiated (5 Gy) and, 30 min later, pulsed with 3 H-thymidine for 20 min. The relative amount of DNA synthesis in irradiated compared with unirradiated cells was determined. **d**, KD-ATM and Flag-tagged S1981A-ATM are dominant-inhibitory for IR-induced intermolecular autophosphorylation of endogenous ATM at serine 1981. The two mutant Flag-tagged ATM constructs or vector alone were transfected into 293T cells. Thirty minutes after exposure to IR (2 Gy), Flag-ATM was cleared from the lysate 2 $\times$ by immunoprecipitation with anti-Flag beads, and the remaining endogenous ATM was immunoprecipitated with anti-ATM antibody. Levels of immunoprecipitated endogenous ATM and levels of contaminating transfected Flag-ATM were determined by anti-ATM and anti-Flag immunoblotting, respectively.

If the serine 1981 domain of ATM (in the unphosphorylated state) binds to the kinase domain of ATM, this could occur either in *cis* or in *trans*. To determine whether ATM protein exists in a protein complex in cells, diploid human fibroblasts were exposed to a range

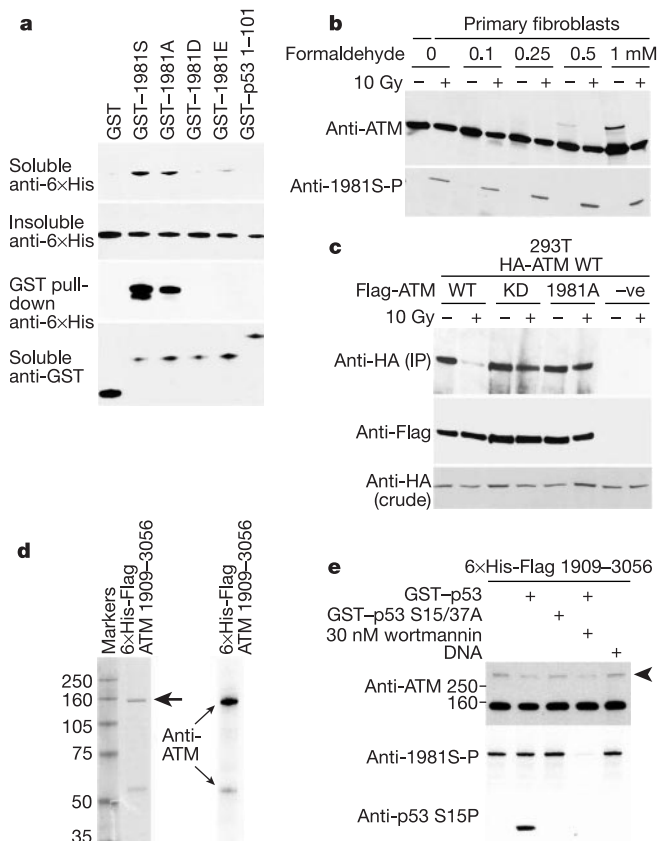


Figure 4 Interaction of ATM domains with proteins. **a**, A recombinant ATM kinase domain is stabilized in *Escherichia coli* by GST fusion proteins containing either 1981S or 1981A. Recombinant 6 $\times$ His-tagged ATM kinase domain and GST alone, GST fusion proteins containing either 1981S, 1981A, 1981D or 1981E, or GST-p53 were expressed in *E. coli* BL21. Whole-cell extracts were prepared, and soluble and insoluble fractions separated. The amount of soluble (upper panel) and insoluble (second panel) recombinant ATM kinase was determined by anti-6 $\times$ His immunoblotting. Binding of GST-1981S and GST-1981A to the kinase domain was shown by co-purification from the soluble fraction using glutathione-agarose (third panel). Expression of GST proteins was assessed by anti-GST immunoblotting (bottom panel). **b**, ATM can be crosslinked into a complex that migrates considerably slower than the ATM monomer before, but not after, exposure to IR. Thirty minutes after irradiation (0 or 10 Gy), exponentially growing fibroblasts were treated with formaldehyde for 10 min. Whole-cell extracts were prepared, and immunoprecipitated ATM was immunoblotted with anti-1981S-P or anti-ATM. **c**, ATM is a dimer before, but not after, exposure to IR. Haemagglutinin (HA)-tagged WT-ATM was co-expressed with either Flag-tagged WT-ATM, KD-ATM or S1981A-ATM. Thirty minutes after irradiation (0 or 10 Gy), ATM immunoprecipitated with anti-Flag was immunoblotted with anti-HA antibody. Total-cell extract was also immunoblotted with anti-HA (bottom panel). **d**, Purified, recombinant ATM fragment (arrow) is soluble in *Saccharomyces cerevisiae*. Coomassie-blue stain of 6 $\times$ His-Flag-ATM 1909–3056 that was affinity-purified against anti-Flag M2 sepharose, eluted with Flag peptide, subjected to size-exclusion chromatography and resolved on a 4–12% denaturing gradient gel. The single additional protein band of approximately 60 kDa was identified as an ATM degradation product by immunoblotting (right panel). **e**, Purified recombinant ATM fragment has kinase activity. The recombinant protein shown in part **d** specifically phosphorylates p53 at serine 15 and itself at serine 1981. Autophosphorylation is not stimulated by the addition of exogenous DNA ends, and both phosphorylation events are inhibited by wortmannin. Anti-ATM immunoblot of the yeast ATM protein fragment reveals a band migrating at a size consistent with a homodimer (top panel, arrowhead).

of formaldehyde concentrations to attempt to covalently crosslink endogenous ATM into the most minimal complex feasible. A prominent complex containing ATM could be immunoprecipitated from the cells following their treatment with 0.5 mM or 1 mM formaldehyde for 10 min (Fig. 4b). This complex migrated electrophoretically considerably more slowly than the denatured ATM monomer, which runs at 370 kDa, and beyond the range of conventional markers of molecular mass. This ATM-containing complex was not present if the cells had been exposed to IR and was not recognized by the anti-1981S-P antibody (Fig. 4b). To establish whether at least two ATM molecules are present in a cellular complex and whether dissociation of the complex is dependent on the phosphorylation of serine 1981, haemagglutinin (HA)-tagged ATM was transfected into 293T cells along with wild-type, kinase-inactive or S1981A Flag-tagged ATM. HA-ATM was immunoprecipitated by anti-Flag sepharose in association with each of the three Flag-tagged ATM proteins (Fig. 4c). After irradiation at 2 Gy, HA-ATM was immunoprecipitated with both kinase-inactive and S1981A Flag-ATM, but was no longer bound to wild-type Flag-ATM. Consistent with the concept that an unphosphorylated serine 1981 is required to bind to the kinase domain and enable it to fold

correctly, ATM constructs with serine 1981 mutated to aspartic acid or glutamic acid could not be expressed in cells (data not shown). Together, these results show that ATM exists as a dimer or a higher-order multimer in unperturbed cells, and that intermolecular ATM autophosphorylation on serine 1981 is required for the dissociation of the complex after DNA damage.

These results provide a mechanistic explanation for the dominant-negative cellular activities of both the kinase-inactive ATM and the S1981A ATM mutant. Because the ATM kinase domain interacts stably with the domain containing the autophosphorylation site, we reasoned that a truncated recombinant ATM molecule that included the phosphorylation and kinase domains might fold correctly and possibly have kinase activity. A plasmid containing 6 × His-Flag-ATM 1923–3056 was expressed in *Saccharomyces cerevisiae*, and a soluble ATM fragment was recovered and purified by anti-Flag affinity and size-exclusion chromatography (Fig. 4d). In an *in vitro* kinase assay, the highly purified ATM fragment had wortmannin-sensitive activity and became autophosphorylated on serine 1981 (Fig. 4e). Further supporting the concept that the ATM protein homodimerizes in cells, some of this ATM fragment migrated at the size of a dimer even under the denaturing conditions of the SDS-PAGE (polyacrylamide gel electrophoresis) gel (Fig. 4e, top panel).

Cellular signals involved in ATM activation

The kinetics, dose responsiveness and stoichiometry of ATM autophosphorylation after IR were examined in primary human fibroblasts. Phosphorylation of serine 1981 was detectable immediately after collecting cells following the 30 s exposure required to deliver 0.5 Gy of IR, was maximal by 5 min, and remained stable and detectable for at least 24 h thereafter (Fig. 5a, b). Dose-dependence studies showed that phosphorylation of serine 1981 was detectable after 0.1 Gy IR, that it was maximal at 0.4 Gy at a 15 min time point, and that no further increase was seen between 1 and 9 Gy (Fig. 5b, d). Exposure to IR at 0.1 Gy should, theoretically, cause just four double-strand breaks in the genomic DNA of a human diploid cell^{26,27}. The extraordinary sensitivity of the detection of DNA strand breaks by ATM phosphorylation was supported by the observation that ATM phosphorylation could also be detected by the introduction of a small number of DNA breaks introduced by a restriction enzyme. ATM phosphorylation was detectable (Fig. 5f) after transfection of the restriction enzyme *I-SceI* into SV-40-transformed fibroblasts that contained two integrated copies of a plasmid containing the *I-SceI* target sequence, a site that has not been found in an unmanipulated mammalian genome²⁸.

To estimate the fraction of cellular ATM protein that becomes phosphorylated after DNA damage, sequential immunoprecipitations of ATM from irradiated primary fibroblasts were carried out with a conventional anti-ATM antibody and with the anti-1981S-P antisera. In the absence of insult, the conventional anti-ATM antibody was able to immunoprecipitate virtually all of the ATM in the first absorption from unirradiated cells, whereas the anti-1981S-P antisera brought down almost no ATM (Fig. 5e, top panel). After exposure to 0.5 Gy IR, the amount of ATM immunoprecipitated by the anti-1981S-P antisera in the first absorption (immunoprecipitation 1, sample 4) was similar to the amount of ATM immunoprecipitated by the conventional anti-ATM antibody (immunoprecipitation 2, sample 3), and was greater than the remaining cellular ATM that was brought down in the second absorption by the conventional anti-ATM antibody (immunoprecipitation 2, sample 4). The results show that well over 50% of the total ATM in an exponentially growing culture of primary human fibroblasts is autophosphorylated on 1981S by 15 min after exposure to 0.5 Gy IR (estimated 18 DNA breaks).

Given that such a high fraction of cellular ATM becomes phosphorylated so rapidly in the presence of so few DNA strand breaks, it seems highly unlikely that the ATM oligomers could require direct binding to DNA strand breaks for activation and

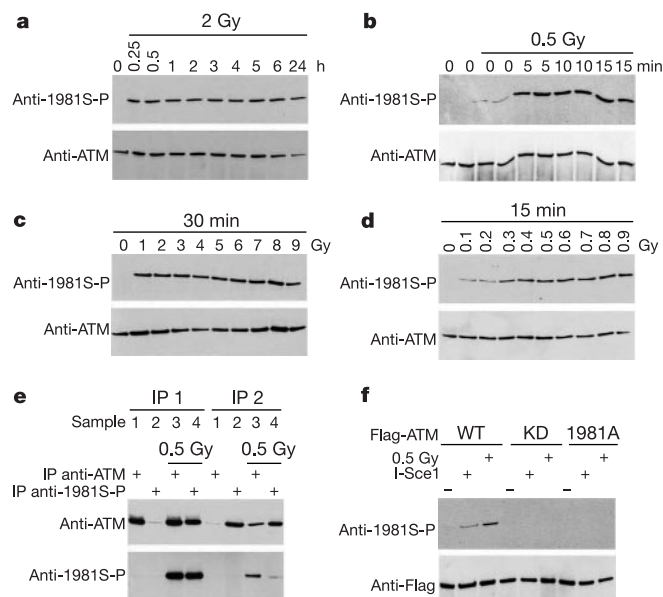


Figure 5 Kinetics, stoichiometry and detection sensitivity of ATM phosphorylation. **a, b**, 1981S phosphorylation is maximal within 5 min after exposure to IR. Exponentially growing primary fibroblasts were exposed to 0.5 or 2 Gy IR. ATM immunoprecipitation was performed at the times indicated, and the levels of 1981S phosphorylation and total immunoprecipitated ATM were assessed by immunoblotting. **c, d**, 1981S phosphorylation is maximal at about 0.4 Gy IR. ATM was immunoprecipitated from exponentially growing primary fibroblasts 30 min after various doses of IR, and the levels of 1981S phosphorylation and total ATM were determined by immunoblotting. **e**, Over 50% of ATM is phosphorylated within 15 min following irradiation at 0.5 Gy. Exponentially growing primary fibroblasts were exposed to 0 or 0.5 Gy IR, and lysates were subjected to sequential immunoprecipitation with conventional anti-ATM or anti-1981S-P antibody followed by conventional anti-ATM antibody. The amount of ATM immunoprecipitated was determined by anti-ATM immunoblotting, and the specificity and completeness of the 1981S-P immunoprecipitation were confirmed by anti-1981S-P immunoblotting. **f**, The anti-1981S-P antibody can detect the presence of a few DNA breaks in the human genome. Constructs expressing Flag-tagged WT-ATM, KD-ATM or S1981A-ATM were co-transfected with a construct encoding the restriction enzyme *I-SceI* into GM00637 cells that had integrated two copies of a plasmid containing the sequence cut by *I-SceI*. As a control, ATM was immunoprecipitated from irradiated cells (0.5 Gy for 30 min). 1981S phosphorylation and total ATM were assessed by immunoblotting with anti-1981S-P or anti-Flag.

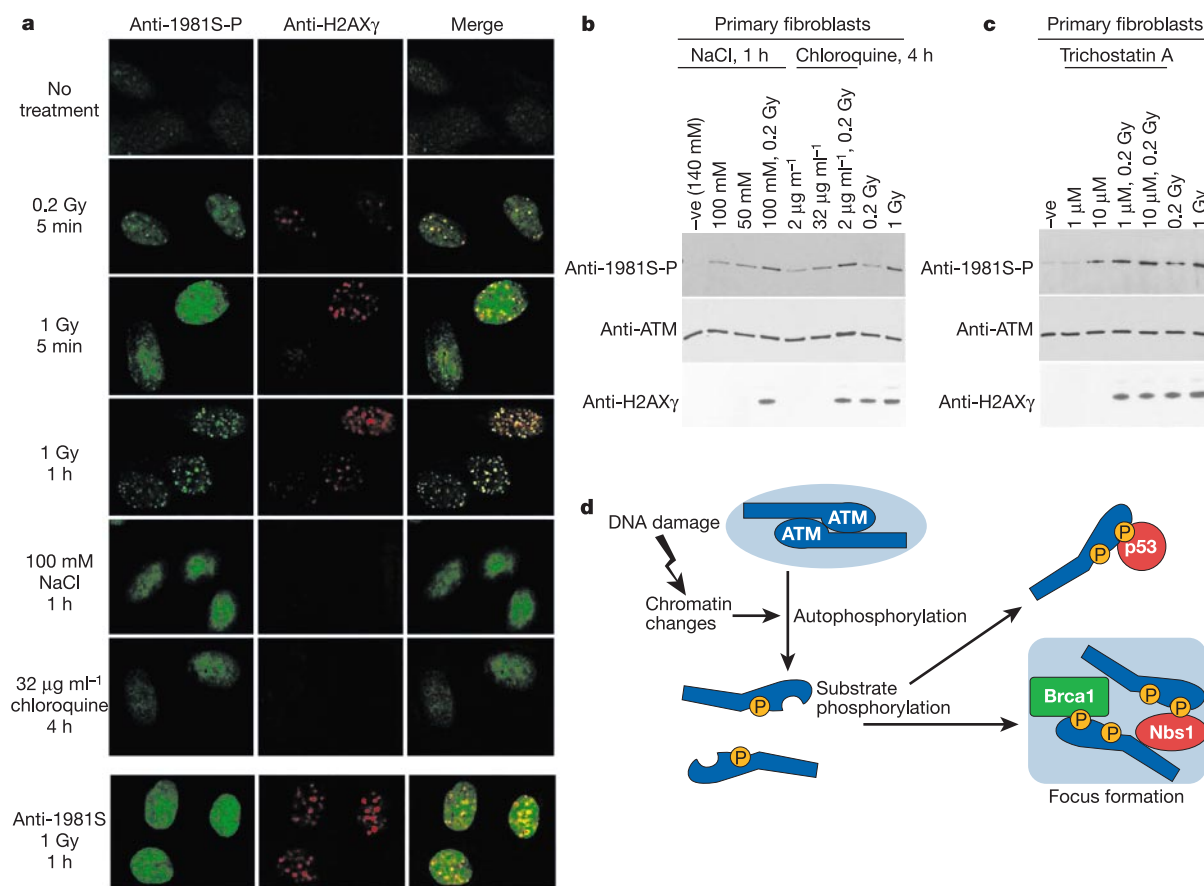


Figure 6 Activation of ATM by chromatin-active agents that do not cause DNA breaks. **a**, Primary fibroblasts grown on glass slides were irradiated (0.2 or 1 Gy) and allowed to recover for 5 min or 1 h, incubated in hypotonic buffer containing 100 mM for 1 h, or treated with chloroquine ($32 \mu\text{g ml}^{-1}$) for 4 h. Cells were fixed, and immunofluorescence for H2AX γ , 1981S-P and 1981S was assessed. **b**, Exponentially growing fibroblasts were incubated in hypotonic (100 mM or 50 mM NaCl) buffer for 1 h or in chloroquine for 4 h. ATM was immunoprecipitated, and the levels of phosphorylated ATM, total ATM and H2AX γ were determined by immunoblotting. No cell death was observed in chloroquine or

in the hypotonic conditions, and all cells recovered when returned to isotonic conditions. **c**, Exponentially growing fibroblasts were incubated in DMEM containing 0.1% FBS for 24 h, and trichostatin was then added for 8 h. ATM was immunoprecipitated and levels of phosphorylated ATM, total ATM and H2AX γ were determined by immunoblotting. **d**, Schematic model of ATM activation after irradiation. DNA strand breaks lead to an alteration of chromatin structures that induce intermolecular autophosphorylations of an ATM dimer on serine 1981, and dissociation of the previously inert dimer. Active ATM monomers are then free to migrate to and phosphorylate substrates such as Nbs1 and p53.

autophosphorylation. These observations suggest that the introduction of DNA strand breaks must cause a change in the nucleus that can activate ATM at a distance from the break itself. As DNA strand breaks introduced by exposure to IR rapidly alter topological constraints on DNA^{29–31}, alterations in some aspect of chromatin structure could meet the criteria of a rapid change that can occur at a distance in the nucleus. Chromatin and chromosome structures can be altered in the absence of DNA breaks by hypotonic conditions^{32,33}, by exposure to chloroquine^{34–36} or by treatment with histone deacetylase inhibitors^{37–39}. Exposure of cells to mildly hypotonic buffers or either of these two chromatin-modifying drugs induced rapid and diffuse phosphorylation of ATM protein, as assessed by immunoblot and immunofluorescence (Fig. 6 and Supplementary Fig. 1).

Because no phosphorylation of histone H2AX or foci of either phosphorylated ATM or H2AX γ were observed with any of these treatments, there is no evidence that any of them caused DNA strand breakage⁴⁰. Interestingly, these chromatin-modifying treatments induce phosphorylation of p53 (Supplementary Fig. 2), whose phosphorylation by ATM does not occur at the site of DNA breaks. By contrast, phosphorylation of both ATM and H2AX were apparent after irradiation, whether it was applied alone or in combination with the other treatments. Furthermore, all three of these agents were able to enhance the amount of ATM

phosphorylation seen after the exposure of cells to submaximal (0.2 Gy) doses of IR (Fig. 6b, c). The patterns of immunofluorescent staining of phosphorylated ATM were also informative. At the earliest time points after IR, the staining was diffuse across the nucleus, but after several minutes, some foci were seen in addition to the diffuse nuclear staining. The diffuse staining, but not the foci, was seen after treatment with the other agents. These patterns are consistent with a diffuse activation of ATM and migration of a fraction of ATM protein to the sites of DNA strand breaks introduced by IR, presumably to phosphorylate substrates at the breaks. Diffuse immunofluorescence, but no ATM foci, was seen when staining the cells with the anti-1981S antibody (Fig. 6a, bottom panel).

Discussion

Characterization of an IR-induced phosphorylation event in ATM and generation of a phosphospecific antibody led to mechanistic insights into the means of signal transduction by the ATM protein kinase; our results point to a new mechanism for regulating the activity of a cellular kinase. We propose that ATM is sequestered in unperturbed cells as a dimer or a higher-order multimer with its kinase domain bound to an internal domain of a neighbouring ATM molecule containing serine 1981 (Fig. 6d). This interdomain interaction must occur for ATM to fold correctly and to remain

stable in the cell; while contained in this complex, ATM is unable to phosphorylate other cellular substrates. After DNA damage, the kinase domain of one ATM molecule phosphorylates serine 1981 of an interacting ATM molecule, and the phosphorylated ATM is then dissociated from the complex and is freed to phosphorylate other substrates in the cell. The kinase-inactive and non-phosphorylatable mutants of ATM retain endogenous ATM in a complex because they cannot be phosphorylated and released after IR, thus inhibiting cellular ATM activity. This mechanism provides an explanation for the dominant-inhibitory property of kinase-inactive ATM, a property that has been used to advantage by many laboratories but has never been understood.

Several different molecular mechanisms have been identified that regulate the activity of protein kinases. Protein kinases are generally restrained in an inactive state with the acquisition of catalytic activity controlled at multiple levels, ranging from the binding of allosteric factors to changes in the subcellular localization of the enzyme⁴¹. Because all protein kinases catalyse the same reaction, their active conformations tend to be structurally similar. However, different classes of kinase have evolved distinct inactive states, and adoption of the catalytic conformation of the enzyme can be impeded in different ways. These include steric hindrance of substrate access to the catalytic domain by an activation loop that is often controlled by phosphorylation⁴²; allosteric regulation of the activation loop via the α C helix, as exemplified by the PSTAIRE helix in the cyclin-dependent kinase family⁴³; pseudosubstrate inhibition of both substrate and nucleotide binding, as seen in twitchin⁴⁴; and intramolecular autoinhibition by N-terminal segments that inhibit catalytic activity, as in the case of the EphB2 receptor kinase⁴⁵. Our observations suggest a new mechanism of kinase activation in which the cellular activity of one ATM kinase molecule is impeded by intermolecular association with an internal domain of a partner ATM molecule. In this model, access of substrates to the catalytic domain is impeded by this association. In some ways, this regulatory model is similar to pseudosubstrate inhibition, with the main variations being that the pseudosubstrate is a domain of itself (albeit in *trans*) and that this partner is not a mimic, but actually becomes a substrate to release the inhibition.

The rapidity, sensitivity and stoichiometry of serine 1981 phosphorylation have ramifications for the mechanisms that must be involved in the initiation of ATM activation. The activation of ATM, as assessed by phosphorylation of this site, is detectable as quickly as cells can be collected following an insult, and is already maximal within 5 min after exposure to an IR dose as low as 0.5 Gy. In addition, phosphorylation of this site is detectable after the introduction of only a few strand breaks in the genome of a human cell, whether the DNA breaks are introduced by low-dose IR or by a restriction enzyme (Fig. 5). Furthermore, activation by the restriction enzyme shows that breaks in the DNA phosphodiester backbone, and not some other cellular effects of IR, are capable of activating ATM. Perhaps the most impressive quantification is that well over 50% of the ATM molecules in the cell become phosphorylated within about 5 min after exposure to 0.5 Gy IR, a dose that would be expected to induce only about 18 double-strand breaks in the genome of a mammalian cell. Although the addition of DNA with free double-stranded ends has been reported to augment the activity of biochemically purified ATM kinase⁴⁶, this enhancement has not been seen when immunoprecipitated endogenous ATM is used in *in vitro* kinase assays²². The relatively high number of ATM molecules that become rapidly phosphorylated and activated, compared with the small number of DNA strand breaks that is required to induce the activation, does not seem consistent with the possibility that ATM has to bind directly to DNA double-strand breaks to become activated; rather, it suggests that the introduction of DNA breaks must somehow signal to ATM molecules at a distance in the cell.

It is conceivable that a protein complex that binds directly to DNA breaks in mammalian cells initiates a signalling pathway that causes ATM activation. However, such a model does not easily explain the impressive rapidity and the magnitude of the ATM phosphorylation reaction that is seen in mammalian cells. An attractive alternative that is consistent with the data presented here is that the introduction of DNA double-strand breaks causes a rapid change in some aspect of higher-order chromatin structure, and that this chromatin alteration initiates ATM activation. DNA in the eukaryotic nucleus is tightly packaged with several levels of organization. DNA is wrapped around nucleosomes that are arranged into ordered arrays of solenoid-like structures. In turn, these are packaged into supercoiled loops, each containing approximately 10^4 – 10^5 base pairs of DNA^{47,48}. IR-induced DNA strand breakage removes topological constraints on DNA loops^{29–31}. The rapidity of ATM activation and the data showing that it can occur at a distance from the DNA strand breaks are consistent with a model in which IR-induced changes in chromatin structure serve as an initiating signal.

The activation of ATM by hypotonic swelling or by treatment with chloroquine or trichostatin, in the absence of detectable DNA strand breaks (Fig. 6), is consistent with this model. Furthermore, it seems that when changes in chromatin structure activate ATM in the absence of DNA strand breaks, ATM substrates that would be phosphorylated at the site of breaks (for example, H2AX) fail to become phosphorylated, whereas substrates present elsewhere in the nucleus (for example, p53) can still be phosphorylated (Fig. 6 and Supplementary Fig. 2). It is tempting to speculate that the amount of ATM phosphorylation becomes maximal at the seemingly low dose of 0.5 Gy (about 18 strand breaks on average per cell), because the number of breaks at this dose can relax the number of higher-order loops that can signal to all of the ATM contained in a cell. Because the activation of ATM is required for most, if not all, appropriate cellular responses to irradiation, these concepts suggest that future investigations of responses to DNA damage should not focus simply on events occurring at the DNA strand break, but rather should include a consideration of more complex nuclear events. □

Methods

Cell culture and immunofluorescence

293T cells, HeLa cells and 1070SK primary human foreskin fibroblasts (HFFs) (less than passage 20, American Type Culture Collection) were cultured in DMEM supplemented with 10% FBS. GM00637 and GM09607 fibroblasts were grown in DMEM containing 15% FBS. GM00536 lymphoblast cells were grown in RPMI supplemented with 10% fetal calf serum. GM00637 stably transfected with the I-SceI restriction enzyme site were obtained from S. Powell (Massachusetts General Hospital). Southern blotting analysis demonstrated two integration sites for the I-SceI plasmid (data not shown). 293T cells were transfected using Eugene (Roche), and HeLa cells, GM00637 and GM09607 with Lipofectamine (Invitrogen). Metabolic labelling was carried out by pre-equilibrating cells in phosphate-free medium for 3 h before the addition of 0.5 mCi ml⁻¹ ³²P-orthophosphate (NEN) for 30 min. Inhibition of DNA synthesis and analysis of the G2/M checkpoint after irradiation were assessed as described previously²⁰. Proteins were crosslinked by incubating cells in formaldehyde/PBS for 10 min at room temperature. Formaldehyde was washed out using PBS containing 100 mM glycine before immunoprecipitation. Hypotonic swelling was carried out for 1 h in PBS containing 0.45% glucose (w/v) and 1% FBS, with the NaCl concentration reduced to either 50 mM or 100 mM. HFFs were incubated with chloroquine in DMEM containing 10% FBS for 4 h. Before an 8 h incubation with trichostatin, HFFs were grown for 24 h in DMEM containing 0.1% FBS. For immunofluorescence experiments, HFFs grown on glass slides were fixed in 50% methanol/50% acetone for 2 h at –20 °C. Cells were incubated with primary antibodies at 1:1,000 (H2AXγ from Upstate Biotechnology) and secondary antibodies at 1:500 (Cy3 anti-mouse and fluorescein isothiocyanate anti-rabbit from Jackson Immunochemicals) in PBS, 10% FBS for 1 h.

Plasmids and recombinant protein purification

Wild-type or mutant Flag-ATM have been described previously⁹. Wild-type Flag-ATM was mutated using the QuikChange site-directed mutagenesis kit (Stratagene). The I-SceI expression plasmid has been described previously²⁰. GST fusion proteins were expressed in BL21 from pGEX 4T1 (Pharmacia). Fusion proteins were purified and GST pull-down experiments carried out by binding to glutathione-sepharose beads (Sigma) in PBS, 0.5% NP-40, and 1 mM 4-(2-aminoethyl)-benzenesulphonyl fluoride hydrochloride (AEBSF)

and 1 mM dithiothreitol. Bound proteins were washed five times in the same buffer and eluted with 20 mM glutathione in 50 mM Tris-HCl (pH 8.0). The ATM kinase domain was expressed from pET28 (Novagen). Recombinant ATM with N-terminal 6 × His and Flag tags was expressed from pYES2 (Invitrogen) in JEL1, a protease-deficient *S. cerevisiae* strain that overexpresses the transcription factor GAL4 driven by the GAL1 promoter⁵⁰. After induction in 2% galactose for 16 h, yeast cells were lysed in 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 0.4 μM aprotinin, 1 mM AEBSF, 1 × soy trypsin inhibitor (Roche), 1.5 mM pepstatin and 42 μM leupeptin by three passages through a French press. Lysates were cleared at 35K in a 45Ti rotor (Beckman). For anti-Flag M2 sepharose (Sigma) affinity purification, Tween 20 and NP-40 were added to 1% and 0.5%, respectively. Flag-ATM was eluted in 100 μg ml⁻¹ Flag peptide (Sigma) in 'buffer A': 50 mM Tris (pH 7.5), 150 mM NaCl, 1% Tween 20, 0.5% NP-40, 50 mM NaF, 1 mM AEBSF and 1 × protease inhibitor mixture from Roche. Size-exclusion chromatography was carried out using a Superdex 200 HR 10/30 column (Pharmacia) in 50 mM sodium phosphate (pH 8.0) containing 150 mM NaCl.

Antibodies and ATM assays

Mammalian cell extracts were prepared in buffer A. Cleared supernatants were immunoprecipitated with anti-Flag M2 sepharose or anti-ATM D1611 (ref. 51), and protein A/G-agarose. Beads were washed twice with buffer A and twice with RIPA buffer. Co-immunoprecipitation was performed in 'buffer B': 50 mM Tris (pH 7.5), 100 μM NaCl, 0.5% Tween 20, 0.2% NP-40, 1 mM AEBSF and 1 × protease inhibitor mixture (Roche). For *in vitro* kinase assays, beads were washed twice with buffer A, twice with buffer A containing 0.5 M LiCl, and twice with kinase buffer: 20 mM HEPES (pH 7.5), 50 mM NaCl, 10 mM MgCl₂ and 10 mM MnCl₂. ATM kinase reactions were carried out at 30 °C for 5 min in 50 μl of kinase buffer containing 10 μCi of [γ-³²P]ATP and 1 μg of GST fusion substrate. Tryptic digestion of Flag ATM immobilized on nitrocellulose, two-dimensional resolution (electrophoresis and chromatography), manual Edman degradation and V8 digestion of peptides isolated from thin-layer chromatography plates were performed as described previously²⁴. Western blotting was performed with anti-Flag M5 (Sigma), anti-ATM MAT3 (a gift of Y. Shiloh), anti-GST (Pharmacia) or anti-6 × His (Sigma). Anti-1981S and anti-1981S-P specific antibodies were generated by immunizing rabbits with the KLFH-conjugated synthetic peptides SLAFEEGQSQTITSS (three animals) and SLAFEEGSpQSQTITSS (six animals) (Rockland Immunochemicals), respectively.

Received 11 October; accepted 16 December 2002; doi:10.1038/nature01368.

- Hartwell, L. H. & Weinert, T. A. Checkpoints: Controls that ensure the order of cell cycle events. *Science* **246**, 629–634 (1989).
- Hartwell, L. H. & Kastan, M. B. Cell cycle control and cancer. *Science* **266**, 1821–1828 (1994).
- Elledge, S. J. Cell cycle checkpoints: Preventing an identity crisis. *Science* **274**, 1664–1672 (1996).
- Kastan, M. B. & Lim, D.-S. The many substrates and functions of ATM. *Nature Rev. Mol. Cell Biol.* **1**, 179–186 (2000).
- Shiloh, Y. & Kastan, M. B. ATM: Genome stability, neuronal development, and cancer cross paths. *Adv. Cancer Res.* **83**, 209–254 (2001).
- Shiloh, Y. Ataxia-telangiectasia and the Nijmegen breakage syndrome: Related disorders but genes apart. *Annu. Rev. Genet.* **31**, 635–662 (1997).
- Savitsky, K. *et al.* A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268**, 1749–1753 (1995).
- Banin, S. *et al.* Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science* **281**, 1674–1677 (1998).
- Canman, C. E. *et al.* Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* **281**, 1677–1679 (1998).
- Maya, R. *et al.* ATM-dependent phosphorylation of Mdm2 on serine 395: Role in p53 activation by DNA damage. *Genes Dev.* **15**, 1067–1077 (2001).
- Matsuoka, S. *et al.* Ataxia telangiectasia-mutated phosphorylates Chk2 *in vivo* and *in vitro*. *Proc. Natl Acad. Sci. USA* **97**, 10389–10394 (2000).
- Chehab, N. H., Malikzay, A., Appel, M. & Halazonetis, T. D. Chk2/hCds1 functions as a DNA damage checkpoint in G₁ by stabilizing p53. *Genes Dev.* **14**, 278–288 (2000).
- Lim, D.-S. *et al.* ATM phosphorylates p95/nbs1 in an S-phase checkpoint pathway. *Nature* **404**, 613–617 (2000).
- Wu, X. *et al.* ATM phosphorylation of Nijmegen breakage syndrome protein is required in a DNA damage response. *Nature* **405**, 477–482 (2000).
- Zhou, B. B. *et al.* Caffeine abolishes the mammalian G₂/M DNA damage checkpoint by inhibiting ataxia-telangiectasia-mutated kinase activity. *J. Biol. Chem.* **275**, 10342–10348 (2000).
- Taniguchi, T. *et al.* Convergence of the Fanconi anemia and ataxia telangiectasia signaling pathways. *Cell* **109**, 459–472 (2002).
- Kim, S. T., Xu, B. & Kastan, M. B. Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. *Genes Dev.* **16**, 560–570 (2002).
- Yazdi, P. T. *et al.* SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. *Genes Dev.* **16**, 571–582 (2002).
- Xu, B., O'Donnell, A. M., Kim, S.-T. & Kastan, M. B. Phosphorylation of serine 1387 in Brca1 is specifically required for the Atm-mediated S-phase checkpoint after ionizing irradiation. *Cancer Res.* **62**, 4588–4591 (2002).
- Xu, B., Kim, S.-T. & Kastan, M. B. Involvement of Brca1 in S-phase and G₂-phase checkpoints after ionizing irradiation. *Mol. Cell. Biol.* **21**, 3445–3450 (2001).
- Bao, S. *et al.* ATR/ATM-mediated phosphorylation of human Rad17 is required for genotoxic stress responses. *Nature* **411**, 969–974 (2001).
- Kim, S.-T., Lim, D.-S., Canman, C. E. & Kastan, M. B. Substrate specificities and identification of putative substrates of ATM kinase family members. *J. Biol. Chem.* **274**, 37538–37543 (1999).

- Sarkaria, J. N. *et al.* Inhibition of phosphoinositide 3-kinase related kinases by the radiosensitizing agent wortmannin. *Cancer Res.* **58**, 4375–4382 (1998).
- Meisenhelder, J., Hunter, T., van der Geer, P. *et al.* in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) Suppl. 48, Ch. 18.9.1–28 (John Wiley & Sons, USA, 1999).
- Bosotti, R., Isacchi, A. & Sonhammer, E. L. FAT: A novel domain in PIK-related kinases. *Trends Biochem. Sci.* **25**, 225–227 (2000).
- Cedervall, B. *et al.* Methods for the quantification of DNA double-strand breaks determined from the distribution of DNA fragment sizes measured by pulsed-field gel electrophoresis. *Radiat. Res.* **143**, 8–16 (1995).
- Ruiz de Almodovar, J. M., Steel, G. G., Whitaker, S. J. & McMillan, T. J. A comparison of methods for calculating DNA double-strand break induction frequency in mammalian cells by pulsed-field gel electrophoresis. *Int. J. Radiat. Biol.* **65**, 641–649 (1994).
- Richardson, C., Elliott, B. & Jasin, M. Chromosomal double-strand breaks introduced in mammalian cells by expression of I-Sce I endonuclease. *Methods Mol. Biol.* **113**, 453–463 (1999).
- Roti Roti, J. L. & Wright, W. D. Visualization of DNA loops in nucleoids from HeLa cells: Assays for DNA damage and repair. *Cytometry* **8**, 461–467 (1987).
- Jaberboansari, A., Nelson, G. B., Roti Roti, J. L. & Wheeler, K. T. Postirradiation alterations of neuronal chromatin structure. *Radiat. Res.* **114**, 94–104 (1988).
- Malyapa, R. S., Wright, W. D., Taylor, Y. C. & Roti Roti, J. L. DNA supercoiling changes and nuclear matrix-associated proteins: Possible role in oncogene-mediated radioresistance. *Int. J. Radiat. Oncol. Biol. Phys.* **35**, 963–973 (1996).
- Earnshaw, W. C. & Laemmli, U. K. Architecture of metaphase chromosomes and chromosome scaffolds. *J. Cell Biol.* **96**, 84–93 (1983).
- Jeppesen, P., Mitchell, A., Turner, B. & Perry, P. Antibodies to defined histone epitopes reveal variations in chromatin conformation and underacetylation of centric heterochromatin in human metaphase chromosomes. *Chromosoma* **101**, 322–332 (1992).
- Krajewski, W. A. Alterations in the internucleosomal DNA helical twist in chromatin of human erythroleukemia cells *in vivo* influences the chromatin higher-order folding. *FEBS Lett.* **361**, 149–152 (1995).
- Jensen, P. B. *et al.* Targeting the cytotoxicity of topoisomerase II-directed epipodophyllotoxins to tumour cells in acidic environments. *Cancer Res.* **54**, 2959–2963 (1994).
- Snyder, R. D. Use of catalytic topoisomerase II inhibitors to probe mechanisms of chemical-induced clastogenicity in Chinese hamster V79 cells. *Environ. Mol. Mutagen.* **35**, 13–21 (2000).
- Krajewski, W. A. Effect of *in vivo* histone hyperacetylation on the state of chromatin fibers. *J. Biomol. Struct. Dyn.* **16**, 1097–1106 (1999).
- Yoshida, M., Horinouchi, S. & Beppu, T. Trichostatin A and trapoxin: Novel chemical probes for the role of histone acetylation in chromatin structure and function. *BioEssays* **17**, 423–430 (1995).
- Kuo, M. H. & Allis, C. D. Role of histone acetylases and deacetylases in gene regulation. *BioEssays* **20**, 615–626 (1998).
- Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858–5868 (1998).
- Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).
- Johnson, L. N. & Noble, M. E. M. Active and inactive protein kinases: Structural basis for regulation. *Cell* **85**, 149–158 (1996).
- De Bondt, H. L. *et al.* Crystal structure of cyclin-dependent kinase 2. *Nature* **363**, 595–602 (1993).
- Hu, S. H. *et al.* Insights into autoregulation from the crystal structure of twitchin kinase. *Nature* **369**, 581–584 (1994).
- Dodelet, V. C. & Pasquale, E. B. Eph receptors and ephrin ligands: Embryogenesis to tumorigenesis. *Oncogene* **19**, 5614–5619 (2000).
- Smith, G. C. M. *et al.* Purification and DNA binding properties of the ataxia-telangiectasia gene product ATM. *Proc. Natl Acad. Sci. USA* **96**, 11134–11139 (1999).
- Vogelstein, B., Pardoll, D. M. & Coffey, D. S. Supercoiled loops and eukaryotic DNA replication. *Cell* **22**, 79–85 (1980).
- Ward, W. S., Partin, A. W. & Coffey, D. S. DNA loop domains in mammalian spermatozoa. *Chromosoma* **98**, 153–159 (1989).
- Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl Acad. Sci. USA* **91**, 6064–6068 (1994).
- Lindsley, J. E. & Wang, J. C. On the coupling between ATP usage and DNA transport by yeast DNA topoisomerase II. *J. Biol. Chem.* **268**, 8096–8104 (1993).
- Alligood, K. J. *et al.* Monoclonal antibodies generated against recombinant ATM support kinase activity. *Hybridoma* **19**, 317–321 (2000).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We acknowledge the technical assistance of H. Davis, M. Reis and D. Woods. We thank S. Powell, Y. Shiloh and R. Abraham for providing reagents, and all members of the Kastan laboratory and J. Cleveland, R. Ivey, T. Izard, P. McKinnon, D. Coffey and C. Sherr for reading the manuscript or for helpful discussions. This work was supported by grants from the National Institutes of Health and by the American Lebanese Syrian Associated Charities of the St Jude Children's Research Hospital.

Competing interests statement The authors declare competing financial interests: details accompany the paper on *Nature's* website (<http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to M.B.K. (e-mail: michael.kastan@stjude.org).

An extrasolar planet that transits the disk of its parent star

Maciej Konacki*, Guillermo Torres†, Saurabh Jha‡‡ & Dimitar D. Sasselov†

* Division of Geological & Planetary Sciences 150-21, California Institute of Technology, Pasadena, California 91125, USA

† Harvard-Smithsonian Center for Astrophysics, 60 Garden St, Cambridge, Massachusetts 02138, USA

‡ Department of Astronomy, University of California, Berkeley, California 94720, USA

Planets orbiting other stars could in principle be found through the periodic dimming of starlight as a planet moves across—or ‘transits’—the line of sight between the observer and the star. Depending on the size of the planet relative to the star, the dimming could reach a few per cent of the apparent brightness of the star. Despite many searches, no transiting planet has been discovered in this way; the one known^{1,2} transiting planet—HD209458b—was first discovered using precise measurements^{2,3} of the parent star’s radial velocity and only subsequently detected photometrically. Here we report radial velocity measurements of the star OGLE-TR-56, which was previously found to exhibit a 1.2-day transit-like light curve^{4,5} in a survey looking for gravitational microlensing events. The velocity changes that we detect correlate with the light curve, from which we conclude that they are probably induced by an object of around 0.9 Jupiter masses in an orbit only 0.023 AU from its star. We estimate the planetary radius to be around 1.3 Jupiter radii and its density to be about 0.5 g cm^{-3} . This object is hotter than any known planet ($\sim 1,900 \text{ K}$), but is still stable against long-term evaporation or tidal disruption.

The advent of high-precision Doppler and timing techniques in the past decade has brought a rich bounty of giant planets^{6–8} as well as smaller, terrestrial-mass pulsar planets⁹. Over one hundred extrasolar giant planets have so far been found by different groups using precise radial velocity measurements⁸. Photometric observations of transiting planets, when combined with radial velocities, yield entirely new diagnostics: the planet size and mean density^{1,2}. Transits supply the orbital inclination and a precise mass for the planet, and in addition they enable a number of follow-up studies^{10–13}. Hence, a large number of transit searches are already planned or underway¹⁴. However, photometry alone cannot distinguish whether the occulting object is a gas giant planet ($\sim 1\text{--}13$ Jupiter masses), a brown dwarf ($\sim 13\text{--}80$ Jupiter masses) or a very late type dwarf star, because such objects have nearly constant radius over a range from around 0.001 to 0.1 solar masses. This critical parameter, the mass of the companion, can be determined from the amplitude of

the radial velocity variation induced in the star.

One of the most successful searches to date is the Optical Gravitational Lensing Experiment (OGLE), which uncovered 59 transiting candidates in three fields in the direction of the Galactic Centre (OGLE-III)^{4,5}, with estimated sizes for the possible companions of about 1–4 Jupiter radii. The large number of relatively faint ($V = 14\text{--}18 \text{ mag}$) candidates to study led to our strategy of a preliminary spectroscopic reconnaissance to detect and reject large-amplitude (high-mass) companions, followed by more precise observations of the very best candidates that remained. Of the 59 OGLE candidates, 20 were unsuitable: one is a duplicate entry, four have no ephemeris (only one transit was recorded), eight show obvious signs in the light curve of a secondary eclipse and/or out-of-eclipse variations (clear indications of a stellar companion), and seven were considered too faint to follow up. We undertook low-resolution spectroscopy of the other 39 candidates in late June and

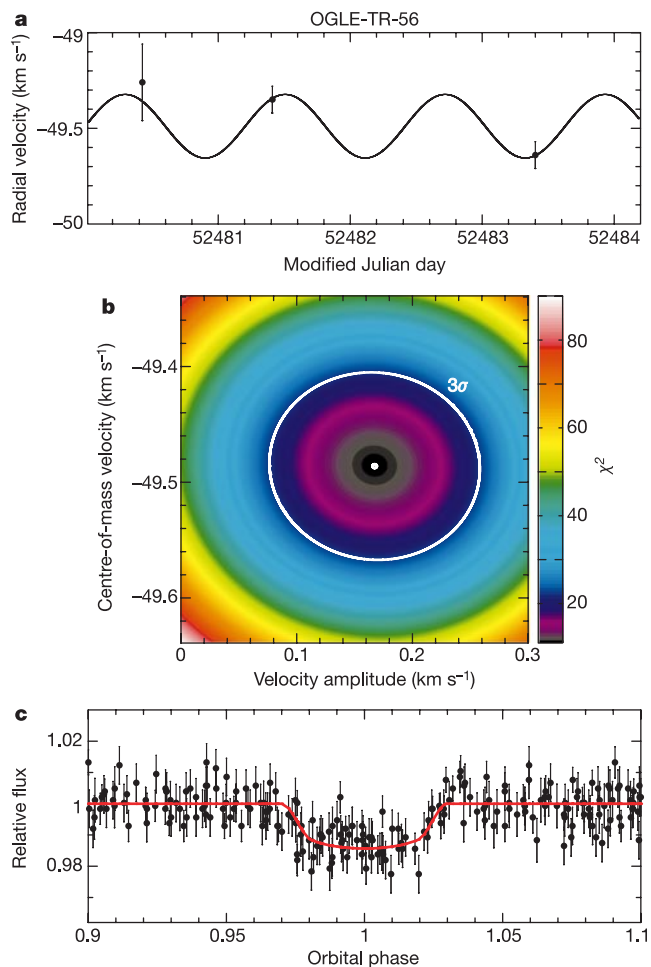


Figure 1 Spectroscopic and photometric observations. **a**, Our radial velocity measurements for OGLE-TR-56 (nightly averages). Note the good agreement with a curve whose only free parameters are the amplitude and systemic velocity; the period and transit epoch are fixed from the OGLE-III photometry. The curve assumes a circularized orbit ($e = 0.0$), as is theoretically expected, and the line thickness corresponds to the phase uncertainty. **b**, χ^2 surface showing the confidence region for our determination of the velocity amplitude and centre-of-mass velocity. The detection of a change in velocity (non-zero velocity amplitude) is formally significant at 6σ , where σ is one standard deviation (see also Table 2). **c**, The photometric transit light curve of OGLE-TR-56 from ref. 5. The transit has an extended flat bottom, and its 1.2%-depth points to a Jupiter-size body (given our determination of a Sun-like primary). The solid red line represents our fitted solution to derive the system parameters. The light curve shows no evidence of other variations or of a secondary eclipse (the hallmark of a strongly blended stellar binary).

Table 1 OGLE-TR-56 radial velocities

Date (MJD)	Radial velocity (km s^{-1})	Error (km s^{-1})
52480.4239	-49.26	0.20
52481.4011	-49.44	0.08
52481.4178	-49.24	0.09
52483.3984	-49.60	0.06
52483.4152	-49.78	0.11

The velocities (reduced to the solar system barycentre) and formal errors are given for each of our individual spectra of OGLE-TR-56. The data indicate a significant variation; a flat line fit gives $\chi^2 \approx 20$ with 4 degrees of freedom (0.06% false-alarm probability), which is considerably worse than a fit to a keplerian orbit model with a fixed ephemeris ($\chi^2 \approx 5$ with 3 degrees of freedom; 17% probability). Having shown, as a check, that the velocities from separate exposures on the same night (originally intended for cosmic ray removal) are not significantly different, we have adopted the nightly averages for subsequent use. A similarly high significance is found for the conclusion that the average velocities are not well fitted by a flat line (99.3% confidence level). MJD, modified Julian day.

mid-July 2002 on the Tillinghast 1.5-m telescope at the F. L. Whipple Observatory (Arizona) and the 6.5-m Magellan I Baade telescope at Las Campanas Observatory (Chile). These spectra were used to eliminate stellar binaries, which can produce shallow, planet-like eclipses caused by blending with light from another star, grazing geometry, or the combination of a large (early-type) primary and a small stellar secondary, but are betrayed by large, easily detected velocity variations (tens of km s^{-1}). We found 25 of the 39 candidates to be stellar binaries, and eight to be of early spectral type. Only six solar-type candidates remained with no detected variations at the few km s^{-1} level (G.T., D.D.S. and S.J., manuscript in preparation).

Subsequently, we used the high-resolution echelle spectrograph (HIRES)¹⁵ on the Keck I 10-m telescope at the W. M. Keck Observatory (Hawaii) on the nights of 24–27 July 2002, to obtain spectra of five of these candidates and measure more precise velocities. OGLE-TR-3 turned out to be the result of grazing eclipses and blending (with even a hint of a secondary eclipse present in the light curve), and the data for OGLE-TR-33, OGLE-TR-10, and OGLE-TR-58 are as yet inconclusive and require further measurements. OGLE-TR-33 exhibits a complex spectral line profile behaviour and could also be a blend. OGLE-TR-10 shows insignificant velocity variation, which is consistent with a sub-Jovian mass planetary companion; OGLE-TR-58 is still inconclusive because of the uncertain ephemeris (M.K., G.T., D.D.S. and S.J., manuscript in preparation). Only OGLE-TR-56 showed clear low-amplitude velocity changes consistent with its 1.21190-day photometric variation⁵, revealing the planetary nature of the companion. With only one bona fide planet (or at most three) among the 39 + 8 objects examined spectroscopically or ruled out on the basis of their light curves, the yield of planets in this particular photometric search has turned out to be very low: at least 94% (possibly up to 98%) of the candidates are ‘false positives’. This is likely to be due in part to the crowded field towards the Galactic centre, which increases the incidence of blends.

We report here our results for OGLE-TR-56 ($I \approx 15.3$ mag). Radial velocities were obtained using exposures of a Th–Ar lamp before and after the stellar exposure for wavelength calibration, and standard cross-correlation against a carefully matched synthetic template spectrum (see Table 1). Our nightly-averaged measurements rule out a constant velocity at the 99.3% confidence level, and are much better represented by a keplerian model of an orbiting planet (Fig. 1a, b). Note that the period and phase of the solid curve are entirely fixed by the transit photometry, as the ephemeris is constrained extremely well by the 12 transits detected so far (A. Udalski, personal communication). The only remaining free parameters are the amplitude of the orbital motion (the key to establishing the mass of the companion) and the centre-of-mass

velocity, both of which can be accurately determined from our velocity measurements. The properties of the planet and the star are summarized in Table 2, and Fig. 1c shows a phased light curve of the transit together with our fitted model.

We performed numerous tests to place limits on any systematic errors in our radial velocities and to examine other possible causes for the variation. These are crucial to assess the reality of our detection. On each night we observed two ‘standards’ (HD209458 and HD179949) which harbour close-in planets with known orbits^{3,16}. We derived radial velocities using the same Th–Ar method as for OGLE-TR-56, and also using the I_2 gas absorption cell to achieve higher accuracy than is possible for our faint OGLE candidates. In Fig. 2 we show that the measured velocity difference between our two standards (HD209458 minus HD179949) is similar for the Th–Ar and I_2 techniques, and more importantly, that both are consistent with the expected velocity change. This indicates that we are able to detect real variations at a level similar to those we see in OGLE-TR-56.

We can rule out the possibility that OGLE-TR-56 is a giant star eclipsed by a smaller main-sequence star, both from a test based on the star’s density inferred directly from the transit light curve¹⁷, and from the very short orbital period. We also examined the spectra for sky/solar spectrum contamination from scattered moonlight; a very small contribution from this source was removed using TODCOR, a two-dimensional cross-correlation technique¹⁸. The separation between the sky lines and the stellar lines is large enough ($\sim 30 \text{ km s}^{-1}$) that the effect on our derived velocities is very small.

Blending of the light with other stars is the most serious concern^{19,20} in a crowded field such as toward the Galactic centre. We have examined the profiles of the stellar spectral lines for asymmetries and any phase-dependent variations that can result from blending. Very little asymmetry is present, and no correlation with phase is observed. In addition, we performed numerical simulations to fit the observed light curve, assuming OGLE-TR-56 is blended with a fainter eclipsing binary. Extensive tests show that with a photometric precision similar to the OGLE data

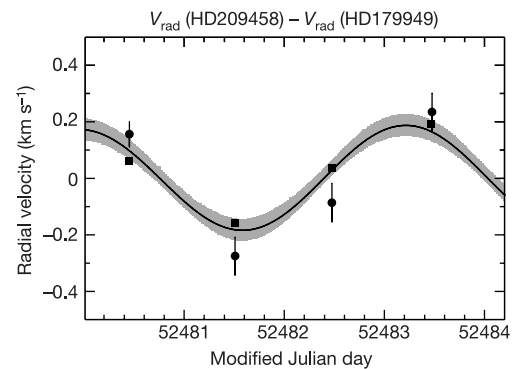


Figure 2 Tests for systematic errors. Predicted radial velocity difference between HD209458 (ref. 3) and HD179949 (ref. 16), our two standard stars with known planets (solid line). We compare these differences with measurements on each of our four observing nights in July 2002. The filled circles are our Th–Ar velocity differences (HD209458 minus HD179949, from the blue echelle orders beyond the iodine spectrum cut-off) with a typical internal uncertainty of about 100 m s^{-1} . These differences should be independent of the wavelength solution itself, and should reveal only the real difference in the Doppler shifts of the stars as well as any systematic problems of an instrumental nature. For comparison, our more precise iodine-cell velocity differences for the same stars (squares) have typical uncertainties of 20 m s^{-1} . The uncertainty introduced by errors in the orbital elements of the standards is indicated by the shaded area. The graph shows that we have succeeded in measuring small changes in velocity on different nights using the standard Th–Ar technique, which indicates the excellent stability of the HIRES instrument. The same technique was applied to our observations of OGLE-TR-56. V_{radial} , radial velocity.

Table 2 Derived stellar and planetary parameters

Parameter	Value
Velocity amplitude	$0.167 \pm 0.027 \text{ km s}^{-1}$
Centre-of-mass velocity	$-49.49 \pm 0.02 \text{ km s}^{-1}$
Orbital period	$1.21190 \pm 0.00001 \text{ days}$
Reference transit epoch (MJD)	52072.185 ± 0.003
Star mass	$(1.04 \pm 0.05) M_{\odot}$
Star radius	$(1.10 \pm 0.10) R_{\odot}$
Limb darkening coefficient (l band)	0.56 ± 0.06
Orbital inclination	$86 \pm 2 \text{ deg}$
Planet distance from star	0.0225 AU
Planet mass	$(0.9 \pm 0.3) M_{\text{J}}$
Planet radius	$(1.30 \pm 0.15) R_{\text{J}}$
Planet density	$0.5 \pm 0.3 \text{ g cm}^{-3}$

The physical properties of the star were derived by modelling the high-resolution spectra with numerical model atmospheres. We find that OGLE-TR-56 is very similar to the Sun, with a temperature of $T_{\text{eff}} \approx 5,900 \text{ K}$. The star’s mass and radius were computed from our stellar evolution tracks²². Combining the stellar parameters with the OGLE-III I -band photometry yields the planetary radius and orbital inclination. The uncertainties shown for the orbital elements are formal errors; the errors for the planet mass and radius additionally reflect our conservative estimate of systematic uncertainties. $M_{\odot}$, $R_{\odot}$, mass and radius of the Sun; M_{J} , R_{J} , mass and radius of Jupiter.

($\sigma \approx 0.003\text{--}0.015$ mag), almost any transit-like light curve can be reproduced as a blend, and only with spectroscopy can these cases be recognized. For each trial simulation, the relative brightness and velocity amplitude of the primary in the eclipsing binary can be predicted. Although a good fit to the photometry of OGLE-TR-56 can indeed be obtained for a model with a single star blended with a fainter system comprising a G star eclipsed by a late M star, the G star would be bright enough that it would introduce strong line asymmetries (which are not seen), or would be detected directly by the presence of a second set of lines in the spectrum. Careful inspection using TODCOR¹⁸ rules this out as well. Therefore, based on the data available, a blend scenario seems extremely unlikely.

This is the faintest ($V \approx 16.6$ mag) and most distant ($\sim 1,500$ pc) star around which a planet with a known orbit has been discovered. The planet is quite similar to the only other extrasolar giant planet with a known radius, HD209458b, except for having an orbit which is almost two times smaller. Thus its substellar hemisphere can heat up to about 1,900 K. However, this is still insufficient to cause appreciable planet evaporation (with a thermal root-mean-square (r.m.s.) velocity for hydrogen of around 7 km s^{-1} compared to a surface escape velocity of around 50 km s^{-1}). The tidal Roche lobe radius of OGLE-TR-56b at its distance from the star is about 2 planet radii. The planet's orbit is most probably circularized ($e = 0.0$) and its rotation tidally locked, but the star's rotation is not synchronized ($v \sin i \approx 3\text{ km s}^{-1}$). Thus the system appears to have adequate long-term stability. Interestingly, OGLE-TR-56b is the first planet found in an orbit much shorter than the current cut-off of close-in giant planets at 3–4-day periods ($\sim 0.04\text{ AU}$)⁸. This might indicate a different mechanism for halting migration in a protoplanetary disk. For example, OGLE-TR-56b may be representative of a very small population of objects—the so-called class II planets, which have lost some of their mass through Roche lobe overflow²¹ but survived in close proximity to the star; a detailed theoretical study of OGLE-TR-56b will be presented elsewhere (D.D.S., manuscript in preparation). These observations clearly show that transit searches provide a useful tool in adding to the great diversity of extrasolar planets being discovered. □

Received 27 November; accepted 30 December 2002; doi:10.1038/nature01379.

- Charbonneau, D., Brown, T. M., Latham, D. W. & Mayor, M. Detection of planetary transits across a Sun-like star. *Astrophys. J.* **529**, L45–L48 (2000).
- Henry, G. W., Marcy, G. W., Butler, R. P. & Vogt, S. S. A transiting “51 Peg-like” planet. *Astrophys. J.* **529**, L41–L44 (2000).
- Mazeh, T. *et al.* The spectroscopic orbit of the planetary companion transiting HD 209458. *Astrophys. J.* **532**, L55–L58 (2000).
- Udalski, A. *et al.* The optical gravitational lensing experiment. Search for planetary and low-luminosity object transits in the Galactic disk. Results of 2001 campaign. *Acta Astron.* **52**, 1–37 (2002).
- Udalski, A. *et al.* The optical gravitational lensing experiment. Search for planetary and low-luminosity object transits in the Galactic disk. Results of 2001 campaign – Supplement. *Acta Astron.* **52**, 115–128 (2002).
- Mayor, M. & Queloz, D. A Jupiter-mass companion to a solar-type star. *Nature* **378**, 355–359 (1995).
- Marcy, G. W. & Butler, R. P. Detection of extrasolar giant planets. *Annu. Rev. Astron. Astrophys.* **36**, 57–98 (1998).
- Schneider, J. *The Extrasolar Planet Encyclopaedia*, available at <http://www.obspm.fr/encycl/encycl.html> (2002).
- Wolszczan, A. & Frail, D. A. A planetary system around the millisecond pulsar PSR1257 + 12. *Nature* **355**, 145–147 (1992).
- Jha, S. *et al.* Multicolor observations of a planetary transit of HD 209458. *Astrophys. J.* **540**, L45–L48 (2000).
- Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W. & Burrows, A. Hubble space telescope time-series photometry of the transiting planet of HD 209458. *Astrophys. J.* **552**, 699–709 (2001).
- Charbonneau, D., Brown, T. M., Noyes, R. W. & Gilliland, R. L. Detection of an extrasolar planet atmosphere. *Astrophys. J.* **568**, 377–384 (2002).
- Brown, T. M., Libbrecht, K. G. & Charbonneau, D. A search for CO absorption in the transmission spectrum of HD 209458b. *Publ. Astron. Soc. Pacif.* **114**, 826–832 (2002).
- Horne, K. in *Proc. Scientific Frontiers of Research on Extrasolar Planets* (eds Deming, D. & Seager, S.) (ASP, San Francisco, in the press).
- Vogt, S. S. *et al.* in *Proc. SPIE Instrum. in Astronomy VIII* (eds Crawford, D. L. & Craine, E. R.) 362–375 (SPIE, Bellingham, WA, 1994).
- Tinney, C. G. *et al.* First results from the Anglo-Australian planet search: A brown dwarf candidate and a 51 Peg-like planet. *Astrophys. J.* **551**, 507–511 (2001).
- Seager, S. & Mallén-Ornelas, G. On the unique solution of planet and star parameters from an extrasolar planet transit light curve. *Astrophys. J.* (submitted); preprint <http://arXiv.org/astro-ph/0206228> (2002).

- Zucker, S. & Mazeh, T. Study of spectroscopic binaries with TODCOR. I: A new two-dimensional correlation algorithm to derive the radial velocities of the two components. *Astrophys. J.* **420**, 806–810 (1994).
- Queloz, D. *et al.* No planet for HD 166435. *Astron. Astrophys.* **379**, 279–287 (2001).
- Santos, N. C. *et al.* The CORALIE survey for southern extra-solar planets. IX. A 1.3-day period brown dwarf disguised as a planet. *Astron. Astrophys.* **392**, 215–229 (2002).
- Trilling, D. E. *et al.* Orbital evolution and migration of giant planets: Modeling extrasolar planets. *Astrophys. J.* **500**, 428–439 (1998).
- Cody, A. M. & Sasselov, D. D. HD 209458: Physical parameters of the parent star and the transiting planet. *Astrophys. J.* **569**, 451–458 (2002).

Acknowledgements We thank A. Udalski and the OGLE team for generous contributions to this project. We are very grateful to S. Kulkarni for support, to R. Noyes and D. Latham for comments, to T. Barlow for assistance with the spectroscopic reductions using the MAKEE Keck Observatory HIRES Data Reduction Software (<http://spider.ipac.caltech.edu/staff/tab/makee>) and to K. Stanek for his encouragement. The data presented here were obtained at the W.M. Keck Observatory (operated by Caltech, the University of California, and NASA), which was made possible by the generous financial support of the W. M. Keck Foundation. M.K. gratefully acknowledges the support of NASA through the Michelson Fellowship programme. G.T. acknowledges support from NASA's Kepler Mission. S.J. thanks the Miller Institute for Basic Research in Science at UC Berkeley for support via a research fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.K. (e-mail: maciej@gps.caltech.edu).

Long-distance teleportation of qubits at telecommunication wavelengths

I. Marcikic*†, H. de Riedmatten*†, W. Tittel*‡, H. Zbinden* & N. Gisin*

* Group of Applied Physics, University of Geneva, CH-1211 Geneva 4, Switzerland

‡ Danish National Research Foundation Center for Quantum Optics (QUANTOP), Institute for Physics and Astronomy, University of Aarhus, Denmark

† These authors contributed equally to this work

Matter and energy cannot be teleported (that is, transferred from one place to another without passing through intermediate locations). However, teleportation of quantum states (the ultimate structure of objects) is possible¹: only the structure is teleported—the matter stays at the source side and must be already present at the final location. Several table-top experiments have used qubits^{2–7} (two-dimensional quantum systems) or continuous variables^{8–10} to demonstrate the principle over short distances. Here we report a long-distance experimental demonstration of probabilistic quantum teleportation. Qubits carried by photons of $1.3\text{ }\mu\text{m}$ wavelength are teleported onto photons of $1.55\text{ }\mu\text{m}$ wavelength from one laboratory to another, separated by 55 m but connected by 2 km of standard telecommunications fibre. The first (and, with foreseeable technologies, the only) application of quantum teleportation is in quantum communication, where it could help to extend quantum cryptography to larger distances^{11–13}.

Since the first article presenting the concept¹ (Fig. 1), quantum teleportation has received much attention. On the conceptual side, it has been proved to be a universal gate for quantum computing¹⁴. In particular, together with quantum memories, it offers the possibility of realizing quantum repeaters with unlimited range¹⁵. But it is fair to say that the fundamental meaning of quantum teleportation for our understanding of quantum nonlocality (and of the structure of space and time) may still be awaiting discovery. On the experimental side, progress in demonstrating the concept has been surprisingly fast. In 1997, two groups—one in Rome, one in Innsbruck—presented results of quantum teleportation using qubits. The Italian group² teleported a qubit carried by one of the

The difficulty is that a complete Bell-state measurement for qubits requires nonlinear optics¹⁸. Hence, one uses linear optics and accepts incomplete measurements, or uses nonlinear optics and accepts very inefficient measurements; or one does not use qubits, or finite dimensional Hilbert spaces. Indeed, it has been shown, first theoretically¹⁹, then experimentally⁸, that the teleportation of continuous variables can, in principle, be fully achieved using linear optics. However, the difficulty is then to produce close to maximally entangled states. This difficulty is even more significant when the distance is increased, because squeezed states of light beams (that is, entangled states for continuous variables) are very vulnerable to losses.

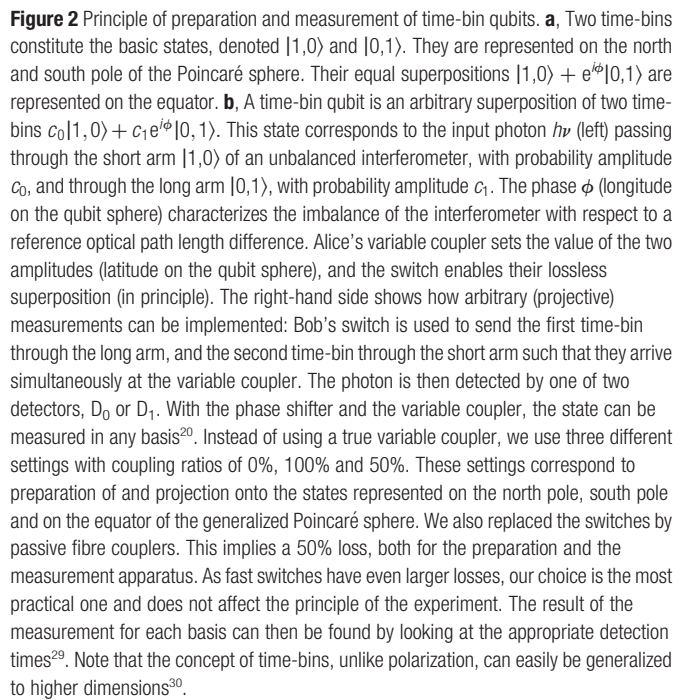
The diagram illustrates a quantum teleportation setup. At the bottom, a "Source of entangled particles" (marked with an asterisk) emits two entangled particles, represented by wavy lines. One particle travels to Alice, and the other travels to Bob. Alice also has an input state $|\psi\rangle$ that she wants to teleport. She performs a "Bell-state measurement" on her two particles. The result of this measurement is sent to Bob via a "Classical channel". Bob then performs a "Unitary transformation" on his particle based on the received classical information, resulting in the teleported state $|\psi\rangle$. A coordinate system with axes t (time) and x (space) is shown in the bottom left corner.

510

$$|\Phi\rangle = \frac{1}{\sqrt{2}}(|1,0\rangle_{\text{Charlie}}|1,0\rangle_{\text{Bob}} + e^{i\varphi}|0,1\rangle_{\text{Charlie}}|0,1\rangle_{\text{Bob}}) \quad (1)$$

The reflected beam is used to create the qubits to be teleported:

where $a_0 = 0, 1$ or $1/\sqrt{2}$, depending on the discrete variable coupler setting, and $a_1 = \sqrt{1 - a_0^2}$. The phase α is defined relative to the reference phase φ . Alice's qubit $|\Psi\rangle_{\text{Alice}}$ is finally sent to Charlie.



the one that projects the two particles onto the singlet entangled state:

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|1,0\rangle_{\text{Alice}}|0,1\rangle_{\text{Charlie}} - |0,1\rangle_{\text{Alice}}|1,0\rangle_{\text{Charlie}}) \quad (3)$$

This takes place when the two photons trigger the detectors labelled C_1 and C_2 in Fig. 3 at times that differ precisely by the time difference between two time-bins. Indeed, each of the two terms in equation (3) may produce this detection result, either with each photon remaining in its fibre or both coupling to the other (hence the π phase shift that corresponds to the minus sign in equation (3)). To achieve this projection, the two photons have to be indistinguishable when they emerge from the beam-splitter²³.

The production of multiple pairs by the entangled photon source should be avoided in many quantum communication protocols²⁴. If the probability of creating a photon pair is the same in both crystals, then, owing to stimulated emission, the probability of creating two pairs in one crystal is the same as the probability of creating one pair in each crystal²⁵. Thus, two times out of three Charlie detects a wrong event²³. In order to detect only the desired events, we

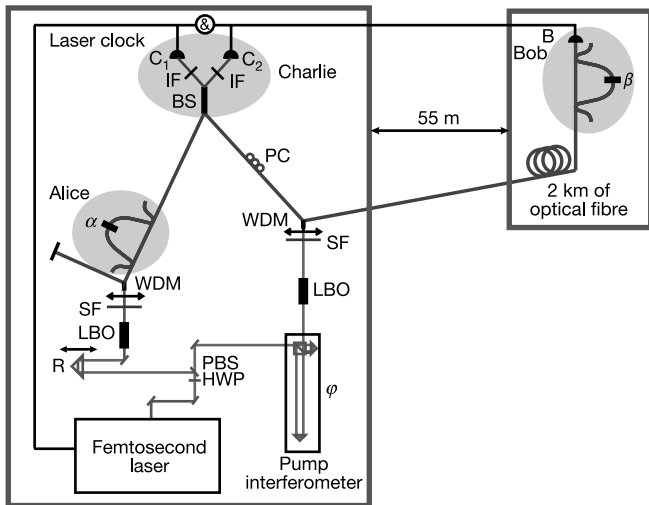


Figure 3 Experimental set-up. Femtosecond laser pulses (with 150-fs pulse width at wavelength $\lambda = 710$ nm, and 76-MHz repetition rate; Coherent Mira 900) are split into two parts using a variable beam-splitter made of a half-wave plate (HWP) and a polarizing beam-splitter (PBS). The reflected beam is used to produce the qubit to be teleported, and the transmitted beam is used to generate the necessary entangled qubit pair. For this purpose the reflected beam is first sent to a nonlinear crystal (lithium triborate, LBO, Crystal Laser), where, by parametric downconversion, a pair of twin photons at telecom communication wavelengths (1.31 and 1.55 μm) is created. Next, the pump light is removed with a silicon filter (SF), and the twin photons are collimated into an optical fibre and separated by a wavelength-division-multiplexer (WDM). The 1.55- μm photon is ignored, and the 1.31- μm photon is sent to Alice, who creates the time-bin qubits to be teleported (with a relative phase α), which she then forwards to Charlie. The retroreflector (R) is used to adjust the arrival time of Alice's qubit at Charlie's. The transmitted beam is first sent through an unbalanced Michelson bulk interferometer (with reference phase $\varphi = 0$) and then through a similar nonlinear crystal, thus producing non-degenerate entangled time-bin qubits. The 1.31- μm photon is sent to Charlie through a polarization controller (PC), and the 1.55- μm photon is sent to Bob who is situated in another laboratory, 55 m away from Charlie and connected by 2 km of optical fibre. Charlie performs a partial Bell-state measurement on Alice's qubit and his part of the entangled pair, using the 50/50 fibre beam-splitter (BS). The photons are detected with single-photon detectors C_1 and C_2 . The interference filters (IF), centred at 1.31 μm with spectral width of 10 nm, are used to render these two particles indistinguishable²³. Finally, whenever the two-particles state is projected onto the $|\Psi^-\rangle$ Bell state, Bob analyses his photon (using an interferometer with relative phase β and a single-photon detector B) to verify that the state encoded by Alice has indeed been teleported. The symbol '&' denotes coincidence detection of the different detector outputs.

decrease the probability of creating entangled qubits relative to the probability of creating the qubit to be teleported. The wrong events are thus reduced to only the cases where two entangled pairs are created; the number of such events can be made arbitrarily small. The ratio of probabilities is controlled by the variable coupler. Eventually, we chose a ratio of 8, with a probability of creating Alice's qubit per laser pulse of around 10%.

Bob is situated in another laboratory, 55 m away from Charlie. To simulate a longer distance, we added 2 km of standard dispersion shifted optical fibre before the teleported photon reaches Bob's analyser. Once Charlie has the information that the (partial) Bell-state measurement was successful, he informs Bob by the classical channel. This operation projects Bob's photon onto the state:

$$|\Psi\rangle_{\text{Bob}} = a_1 e^{i\alpha} |1,0\rangle_{\text{Bob}} - a_0 |0,1\rangle_{\text{Bob}} \quad (4)$$

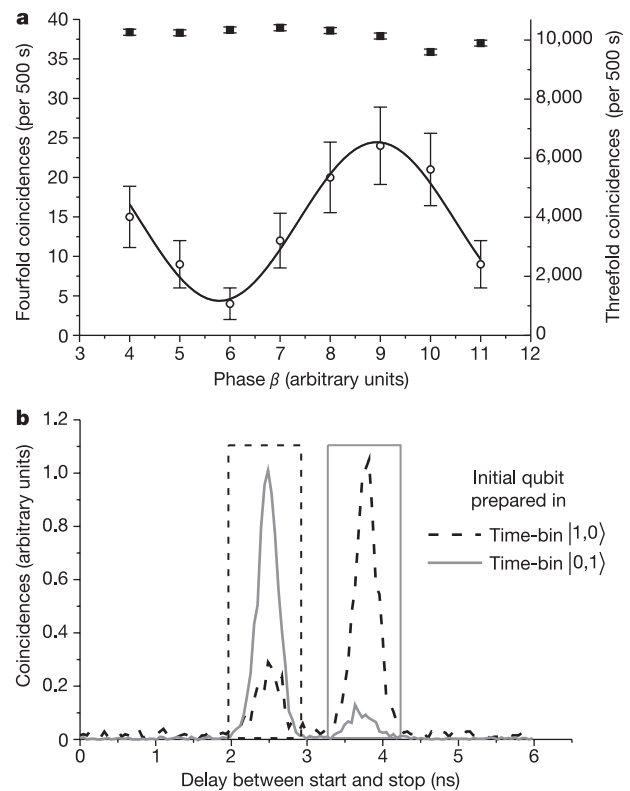


Figure 4 Experimental results. **a**, Teleportation of a qubit consisting of a coherent superposition of two time-bins (equatorial states). Open circles represent Bob's count rate conditioned on the projection on the $|\Psi^-\rangle$ Bell state (fourfold coincidences) as a function of the phase β in his interferometer. The clear interference pattern with visibility $V = (70 \pm 5)\%$ confirms that teleportation takes place with a fidelity of $(85 \pm 2.5)\%$. Squares represent the three-fold coincidence count rates between laser pulse, one of Charlie's detectors (C_1) and Bob's detector. They illustrate the stability of the set-up. **b**, Teleportation of qubits $|1,0\rangle$ (north pole, dashed curve) and $|0,1\rangle$ (south pole, plain curve). The corresponding fidelities are $(77 \pm 3)\%$ and $(88 \pm 3)\%$, respectively. As indicated in equation (4), the projection onto the $|\Psi^-\rangle$ Bell state leads to a bit-flip of the teleported state. The different results for the two input states are due to the fact that in our experimental set-up the detection of the first photon, in mode C_1 , triggers the two other detectors, C_2 and B. When we prepare the state $|1,0\rangle$ the first detection is mostly due to Alice's photon, because as explained above, we produce eight times more qubits to be teleported than entangled pairs. Hence, the two detectors C_2 and B are often triggered without any photon present, leading to an increasing number of accidental coincidences, that is, wrong events. When preparing the other state $|0,1\rangle$, the first detection in mode C_1 can only be due to a photon coming from the source that generates entangled pairs, or to a dark count. As these events occur much less frequently than in the first-mentioned case, the corresponding teleportation fidelity is less affected by accidental coincidences.

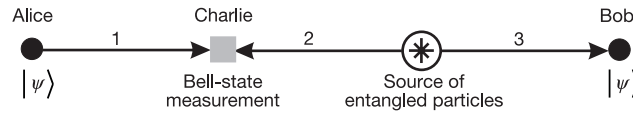


Figure 5 Quantum teleportation used as a quantum relay. Quantum relays extend the range of quantum cryptography from tens of kilometres to hundreds of kilometres, though not to unlimited ranges. The basic idea is as follows^{11–13}: in quantum cryptography, the noise is dominated by the detector dark counts; hence the noise is almost independent of the distance. The signal, however, decreases exponentially with distance because of the attenuation. With realistic numbers, this sets a limit close to 80 km. But if it were possible to verify at some points along the quantum channel whether or not the photon is still there, Bob could refrain from opening the detector when there is no photon. This simple idea is impractical, because it requires (at present

unrealistic) photon number quantum non-demolition measurements³¹. However, consider a channel divided into sections—for example, the three sections as illustrated here. Assume that the photon sent by Alice down the first section is teleported to Bob using the entangled photon pair generated between sections 2 and 3. Two photons travel towards Bob, and one towards Alice. But, because the time ordering of the measurements is irrelevant⁷, the logical qubit can be considered to propagate all the way from Alice to Bob. Accordingly, the Bell-state measurement and the two-photon source act on this logical qubit as a non-demolition measurement

In order to recover Alice's qubit state (equation (2)), Bob should apply the σ_y unitary transformation, consisting of a bit flip ($|1,0\rangle \leftrightarrow |0,1\rangle$) and a phase flip (of relative phase π). However, these unitary operations are not necessary to prove that teleportation takes place.

To show that our teleportation set-up operates correctly, Bob analyses the received photon with an analyser adapted for the wavelength of 1.55 μm (ref. 24; see Fig. 2). The analysis basis thus contains the vector:

$$k_0|1,0\rangle + k_1e^{i\beta}|0,1\rangle \quad (5)$$

where $k_0 = 0,1$ or $1/\sqrt{2}$, depending on the variable coupler setting, and $k_1 = \sqrt{1 - k_0^2}$.

One of Charlie's photons is detected by the passively quenched germanium avalanche photo diode (APD) C_1 working at liquid nitrogen temperature in the so-called Geiger mode²⁶ (quantum efficiency $\eta = 10\%$, dark count rate d.c.r. = 35 kHz, from NEC). To reduce the noise we make a coincidence between the Germanium APD and a trigger from the laser pulses. The other photon arriving at Charlie's and Bob's photon are detected with Peltier-cooled InGaAs APDs (C_2 and B, respectively) working in the so-called gated mode²⁷ ($\eta = 30\%$, d.c.r. = 10^{-4} per ns, from id Quantique). The trigger is given by the coincidence between the germanium APD and the laser pulse. Finally, the signals of the APDs are sent to fast coincidence electronics. We monitor fourfold coincidences with a time-to-amplitude converter, where the start is given by a successful Bell-state measurement (detectors C_1 and C_2 + laser pulse) and the stop by Bob's detector B. The start plays the role of the classical information that Charlie sends to Bob. We choose to record only the events when the photon in output C_2 arrives with a time difference $\Delta\tau$ after the photon in output C_1 . We also record the coincidence between detectors C_1 and B. The rate should remain constant, as it contains no information about the Bell-state measurement result. This provides a control of the stability of the entire set-up.

In order to show that our teleportation set-up is universal, we report the teleportation of two different classes of states. The first class is composed of superposition of two time-bins, and hence corresponds to state represented by points on the equator of the generalized Poincaré sphere (Fig. 2). The second class consists of the two time-bins themselves, represented by the north and south poles of the sphere. The quality of the teleportation is usually reported in terms of fidelity $\bar{F}$ that is, the probability that Bob's qubit, described by the density matrix ρ_{out} successfully passes an analyser testing that it is indeed in the state Ψ_{Alice} prepared by Alice, averaged over all possible Ψ_{Alice} :

$$\bar{F} = \int \langle \Psi_{\text{Alice}} | \rho_{\text{out}} | \Psi_{\text{Alice}} \rangle d\Psi_{\text{Alice}} \quad (6)$$

The linearity of quantum mechanics implies that

$$\bar{F} = \frac{2}{3}F_{\text{equator}} + \frac{1}{3}F_{\text{poles}} \quad (7)$$

where F_{equator} and F_{poles} are the averaged fidelities for the equatorial and pole states, respectively.

To measure the teleportation fidelity F_{equator} of the equatorial states we scanned the phase β in Bob's interferometer. This results in the normalized coincidence count rate:

$$R_C = \frac{1 - V \cos(\alpha + \beta)}{2} \quad (8)$$

corresponding to Bob's state $\rho_{\text{out}} = V|\Psi_{\text{Alice}}\rangle\langle\Psi_{\text{Alice}}| + (1 - V)(1/2)$, where V is the visibility, which can theoretically reach the value of 1. Accordingly, the fidelity equals 1 with probability V and equals $1/2$ with probability $1 - V$, hence $F_{\text{equator}} = (1 + V)/2$. Fig. 4a shows a good result leading to a fidelity of $(85 \pm 2.5)\%$. By performing repeatedly many experiments over a few weeks with different phases α , we typically obtain fidelities around $(80.5 \pm 2.5)\%$.

The preparation of the two other states, represented by the north and south poles, implies the use of a variable coupler at settings of 0% and 100%, respectively. We realize this by using two different fibres of appropriate lengths. For the measurement, Bob uses only one fibre and looks for detections at appropriate times. As shown in equation (4), when Alice sends such a state, Bob receives the orthogonal state, that is, he should bit-flip his qubit to recover Alice's state. The corresponding fidelity is the probability of detecting the right state when measuring in the north–south basis, $F_{\text{poles}} = R_{\text{correct}}/(R_{\text{correct}} + R_{\text{wrong}})$ (Fig. 4b). The measured fidelity for the $|1,0\rangle$ input state is $(77 \pm 3)\%$ and for the $|0,1\rangle$ input state $(88 \pm 3)\%$. Accordingly the mean value is $F_{\text{poles}} = (82.5 \pm 3)\%$.

From these results, we conclude that the overall mean fidelity is $\bar{F} = (81.2 \pm 2.5)\%$ (equation (7)). This value is six standard deviations above the maximum fidelity of 66.7% achievable with the best protocol using no entanglement²⁸.

The difference between our experimental results and the ideal theoretical case could be due to various imperfections. First, our Bell-state measurement is not perfect; although we tried to detect only the events when there is only one photon in each mode, there is still an 11% chance that we make a spurious coincidence. The value can be found in the noise of the teleportation of the $|0,1\rangle$ input state, which is essentially due to the creation of double entangled pairs. The fidelity might also be reduced owing to different polarization or spectra of the two photons when arriving at the beam-splitter, or to remaining temporal distinguishability. Second, the creation and analysis of the qubits is not perfect²⁴. Third, detector dark counts also reduce the measured fidelity. Last, one notes that teleportation occurs only when there is a projection onto the $|\Psi^-\rangle$ state, and that only one-eighth of the qubits sent by Alice are teleported: this renders our realization probabilistic, even assuming perfect detectors. This is a drawback from a fundamental point of view. However, if

quantum teleportation were used as a quantum relay (see Fig. 5) in quantum cryptography, then the probabilistic nature of our teleportation scheme would only affect the count rate, not the quality of the quantum relay. Our scheme would be useful for this application. □

Received 19 September; accepted 13 December 2002; doi:10.1038/nature01376.

1. Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
2. Boschi, D., Branca, S., De Martini, F., Hardy, L. & Popescu, S. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
3. Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
4. Nielsen, M. A., Knill, E. & Laflamme, R. Complete quantum teleportation using nuclear magnetic resonance. *Nature* **396**, 52–55 (1998).
5. Kim, Y.-H., Kulik, S. P. & Shih, Y. Quantum teleportation of polarization state with a complete Bell state measurement. *Phys. Rev. Lett.* **86**, 1370–1373 (2001).
6. Lombardi, E., Sciarrino, F., Popescu, S. & De Martini, F. Teleportation of a vacuum-one-photon qubit. *Phys. Rev. Lett.* **88**, 070402 (2002).
7. Jennewein, T., Weihs, G., Pan, J.-W. & Zeilinger, A. Experimental nonlocality proof of quantum teleportation and entanglement swapping. *Phys. Rev. Lett.* **88**, 017903 (2002).
8. Furusawa, A. *et al.* Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
9. Babichev, S. A., Ries, J., Lvovsky, A. I. Quantum scissors: teleportation of single-mode optical states by means of a nonlocal single photon. Preprint quant-ph/0208066 at (<http://xxx.lanl.gov>) (2002).
10. Bowen, W. P. *et al.* Experimental investigation of continuous variable quantum teleportation. Preprint quant-ph/0207179 at (<http://xxx.lanl.gov>) (2002).
11. Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195 (2002).
12. Waks, E., Zeevi, A. & Yamamoto, Y. Security of quantum key distribution with entangled photons against individual attacks. *Phys. Rev. A* **65**, 052310 (2002).
13. Jacobs, B. C., Pittman, T. B. & Franson, J. D. Quantum relays and noise suppression using linear optics. *Phys. Rev. A* **66**, 052307 (2002).
14. Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).
15. Briegel, H.-J., Dur, W., Cirac, J. I. & Zoller, P. Quantum repeaters: The role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
16. Zukowski, M., Zeilinger, A., Horne, M. A. & Ekert, A. K. “Event-ready-detectors” Bell experiment via entanglement swapping. *Phys. Rev. Lett.* **71**, 4287–4290 (1993).
17. Pan, J.-W., Bouwmeester, D., Weinfurter, H. & Zeilinger, A. Experimental entanglement swapping: Entangling photons that never interacted. *Phys. Rev. Lett.* **80**, 3891–3894 (1998).
18. Lütkenhaus, N., Calsamiglia, J. & Suominen, K.-A. Bell measurements for teleportation. *Phys. Rev. A* **59**, 3295–3300 (1999).
19. Braunstein, S. L. & Kimble, H. J. Teleportation of continuous quantum variables. *Phys. Rev. Lett.* **80**, 869–872 (1998).
20. Tittel, W. & Weihs, G. Photonic entanglement for fundamental tests and quantum communication. *Quant. Inf. Comput.* **1**, 3–56 (2001).
21. Brendel, J., Tittel, W., Zbinden, H. & Gisin, N. Pulsed energy-time entangled twin-photon source for quantum communication. *Phys. Rev. Lett.* **82**, 2594–2597 (1999).
22. Thew, R. T., Tanzilli, S., Tittel, W., Zbinden, H. & Gisin, N. Experimental investigation of the robustness of partially entangled photons over 11 km. *Phys. Rev. A* **66**, 062304 (2002).
23. De Riedmatten, H., Marcikic, I., Tittel, W., Zbinden, H. & Gisin, N. Quantum interferences with photon pairs created in spatially separated sources. *Phys. Rev. A* (in the press); preprint quant-ph/0208174 at (<http://xxx.lanl.gov>) (2002).
24. Marcikic, I., De Riedmatten, H., Tittel, W., Zbinden, H. & Gisin, N. Femtosecond time-bin entangled qubits for quantum communication. *Phys. Rev. A* **66**, 062308 (2002).
25. Lamas-Linares, A., Howell, J. C. & Bouwmeester, D. Stimulated emission of polarization-entangled photons. *Nature* **412**, 887–890 (2001).
26. Owens, P. C. M., Rarity, J. G., Tapster, P. R., Knight, D. & Townsend, P. D. Photon counting with passively quenched germanium avalanche. *Appl. Opt.* **33**, 6895–6901 (1994).
27. Stucki, D., Ribordy, G., Stefanov, A. & Zbinden, H. Photon counting for quantum key distribution with Peltier cooled InGaAs/InP APD's. *J. Mod. Opt.* **48**, 1967–1981 (2001).
28. Massar, S. & Popescu, S. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.* **74**, 1259–1263 (1995).
29. Tittel, W., Brendel, J., Zbinden, H. & Gisin, N. Quantum cryptography using entangled photons in energy-time Bell states. *Phys. Rev. Lett.* **84**, 4737–4740 (2000).
30. De Riedmatten, H., Marcikic, I., Zbinden, H. & Gisin, N. Creating high dimensional entanglement using mode-locked laser. *Quant. Inf. Comput.* **2**, 425–433 (2002).
31. Grangier, P., Levenson, J. A. & Poizat, J.-P. Quantum non-demolition measurements in optics. *Nature* **396**, 537–542 (1998).

Acknowledgements We thank M. Legré for discussions, and C. Barreiro and J.-D. Gautier for technical support. Financial support by the Swiss OFES and NSF within the framework of the European IST project Qucomm and the Swiss National Center for Quantum Photonics is acknowledged. W.T. acknowledges support from the ESF Programme Quantum Information Theory and Quantum Computation (QIT).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.G. (e-mail: Nicolas.Gisin@Physics.Unige.ch).

A colloidal model system with an interaction tunable from hard sphere to soft and dipolar

Anand Yethiraj*† & Alfons van Blaaderen*

* Soft Condensed Matter, Debye Institute, Utrecht University, Padualaan 5, 3584CC Utrecht, and FOM Institute for Atomic and Molecular Physics, Kruislaan 407, 1098 SJ Amsterdam, The Netherlands

Monodisperse colloidal suspensions of micrometre-sized spheres are playing an increasingly important role as model systems to study, in real space, a variety of phenomena in condensed matter physics—such as glass transitions and crystal nucleation^{1–4}. But to date, no quantitative real-space studies have been performed on crystal melting, or have investigated systems with long-range repulsive potentials. Here we demonstrate a charge- and sterically stabilized colloidal suspension—poly(methyl methacrylate) spheres in a mixture of cycloheptyl (or cyclohexyl) bromide and decalin—where both the repulsive range and the anisotropy of the interparticle interaction potential can be controlled. This combination of two independent tuning parameters gives rise to a rich phase behaviour, with several unusual colloidal (liquid) crystalline phases, which we explore in real space by confocal microscopy. The softness of the interaction is tuned in this colloidal suspension by varying the solvent salt concentration; the anisotropic (dipolar) contribution to the interaction potential can be independently controlled with an external electric field ranging from a small perturbation to the point where it completely determines the phase behaviour. We also demonstrate that the electric field can be used as a pseudo-thermodynamic temperature switch to enable real-space studies of melting transitions. We expect studies of this colloidal model system to contribute to our understanding of, for example, electro- and magneto-rheological fluids.

Recent advances in quantitative three-dimensional (3D) real-space analysis of the structure and dynamics of colloids have been accompanied by progress in the synthesis of model spheres with a controllable size and surface chemistry enabling the manipulation of the interaction potentials between the spheres^{5,6}. Together with the well developed scattering techniques for studying colloidal systems^{7–12}, this has led to new insights into the glass transition of simple glass formers^{2,3,12,13}, crystal nucleation and growth^{4,7–10,14}, and the role of the attractive part of interaction potentials on phase behaviour¹¹.

The simplest colloidal model system with short-ranged repulsive interactions is either a sterically stabilized particle suspension^{5,14,15}, or an aqueous suspension of micrometre-sized latex spheres with salt added to ensure that the inverse Debye screening length, $1/\kappa$, a measure for the range of the repulsion, is short; $\kappa R = 33$ for spheres of radius $R = 1 \mu\text{m}$ in water at a monovalent salt concentration of $c = 0.1 \text{ mM}$. To realize almost hard-sphere behaviour, the always-present van der Waals attractions have to be reduced by matching the index of refraction in the visible. This allows the interactions to be tuned to almost hard-sphere-like even for micrometre-sized colloids that can be quantitatively imaged in 3D by confocal microscopy. The role of gravity (far more important in colloids than in the crystallization of atomic and molecular fluids) is minimized by matching the density of the spheres and solvent. A step-up in complexity could involve either introducing a dipolar interaction, thus making the interaction anisotropic, or extending

† Present address: Department of Chemistry, 2036 Main Mall, Vancouver, British Columbia, Canada V6T 1Z1.

the range of the repulsion (for example, increasing the screening length $1/\kappa$ such that $\kappa R \approx 1$). The latter can be achieved in aqueous suspensions by removing ions by using ion-exchange resin, but it is difficult to reduce concentrations below $1 \mu\text{M}$ (that is, $1/\kappa \approx 300 \text{ nm}$). We will show that in less polar organic solvents, screening lengths of $1/\kappa \approx 12 \mu\text{m}$ can be achieved. Dipolar interactions are induced, and finely controlled via an external electric field (the 'electrorheological effect'¹⁶). Studies of structure evolution in brownian hard-sphere fluids with an added dipolar interaction^{17–20} show the formation of strings, then sheets and finally 3D crystallites. As far as we know, long-range repulsive colloidal spheres with anisotropic interactions have not yet been studied.

Here we present confocal microscopy studies of charged and sterically stabilized polymethyl methacrylate (PMMA) spheres (radius $R = 1.0$ and $2.0 \mu\text{m}$) in a density- and refractive-index-matched organic solvent mixture; this mixture polar enough (dielectric constant, $\epsilon = 5$ – 6) to have electric-field-induced dipolar interactions. Two solvent mixtures were used: cycloheptyl bromide/*cis*-decalin and cyclohexyl bromide/*cis*-decalin. The PMMA spheres were monodisperse, fluorescently labelled and sterically stabilized^{5,21}. Three-dimensional particle coordinates of the colloids were obtained as described before^{2,3}. Charge on PMMA spheres in similar solvent mixtures to those used here was previously reported to result from the presence of water or from the dye incorporated in the particles^{4,22}. However, we found that neither drying and distilling the cycloheptyl bromide, nor the use of undyed particles, had an

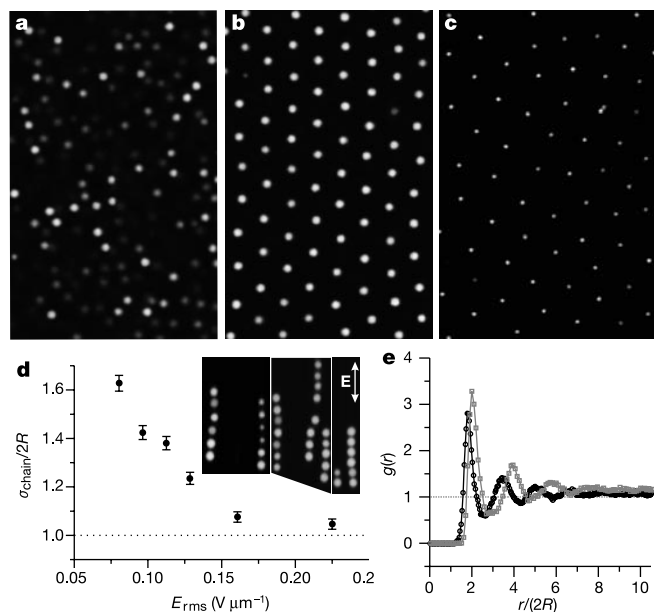


Figure 1 Tunability of interactions. Spheres, radius $R = 2.0 \mu\text{m}$; suspension medium, cyclohexyl bromide-*cis*-decalin. **a**, Fluid at volume fraction $\varphi = 0.02$, salt concentration $c = 40 \mu\text{M}$, maximum of first peak in 2D pair correlation function $g(r)$, $8.0 \mu\text{m}$. **b**, Crystalline (f.c.c.) at $\varphi = 0.02$ and no salt, interparticle spacing, $11.7 \mu\text{m}$. **a, b**, Confocal micrographs of 2D snapshots $50 \mu\text{m}$ into the sample bulk. **c**, $\varphi = 0.002$ (no salt). A plane $33 \mu\text{m}$ above the bottom surface is shown: the interparticle spacing $\sigma \approx 26.7 \mu\text{m}$. **d**, The string interparticle spacing σ_{chain} decreases with increasing field, owing to a balance between anisotropic attraction and charge repulsion. Inset, a 1-MHz a.c. external field $E_{\text{rms}} = 0.09, 0.14$ and $0.25 \text{ V } \mu\text{m}^{-1}$ respectively; field direction, z as shown) induces the formation of strings. At fields above $E_{\text{rms}} = 0.35 \text{ V } \mu\text{m}^{-1}$, the strings span the sample so there is 1D periodicity. **e**, In the x - y plane, however, the suspension remains fluid-like. Shown is 2D $g(r)$ at zero-field (black circles) and $E_{\text{rms}} = 0.35 \text{ V } \mu\text{m}^{-1}$ (grey squares). The structure in both cases is liquid-like: however the peaks in the 'string fluid' are shifted farther out, and higher due to the in-plane registry caused by the 1D periodicity.

effect on the softness of the potential. We found that we could control the range of the repulsive interactions by adding the salt tetrabutylammonium chloride.

Hard-sphere suspensions are fluid at low particle volume fractions φ but crystallize into a close-packed crystal structure for $\varphi > 0.545$. A random hexagonal close packing (r.h.c.p.) is observed experimentally, as opposed to the theoretically predicted face-centred cubic (f.c.c.) crystal²³. When the potential is made soft (κR smaller), crystallization occurs for smaller φ . The crystal structure is then f.c.c. although for a range of κR values, there can be an intervening body-centred cubic (b.c.c.) phase. The phase behaviour has been extensively studied^{9,24–26}, and modelled (and reproduced, with a renormalized surface charge density²⁶) by a pure Yukawa²⁴ or a hard-sphere-plus-Yukawa²⁷ potential. In Fig. 1a–c, we demonstrate the fine control of the softness of the interparticle interaction. Samples in Fig. 1a and b were prepared at identical volume fraction $\varphi = 0.02$. The sample in Fig. 1a ($c = 40 \mu\text{M}$) was fluid, while that in Fig. 1b, in a pure suspension without added salt (corresponding to $1/\kappa \approx 12 \mu\text{m}$) was crystalline. The system shown in Fig. 1c, also in the pure suspension, was crystalline at $\varphi = 0.002$. At $c = 260 \mu\text{M}$ the inter-particle spacing was $\sigma/2R = 1.16 \pm 0.01$,

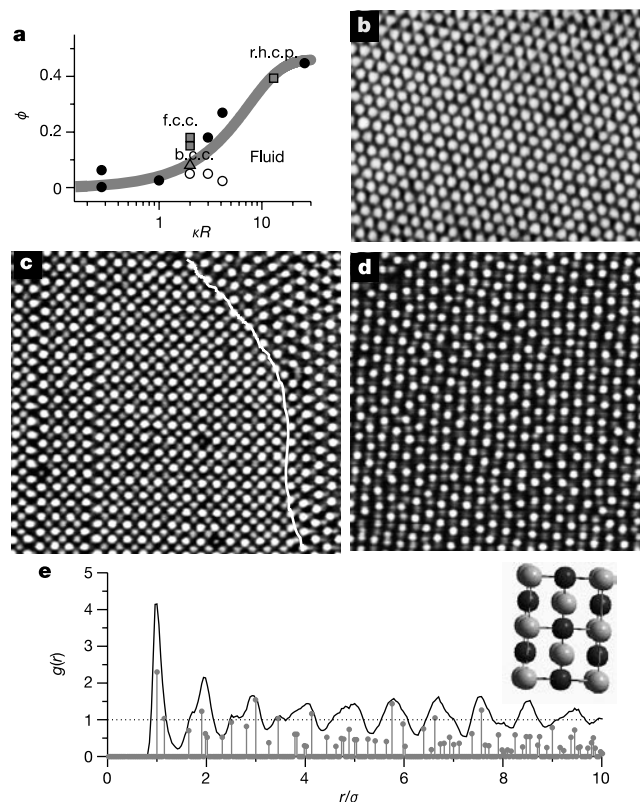


Figure 2 Influence of softness on the phase behaviour. **a**, Phase diagram as function of softness, κR ($1/\kappa$ is the inverse Debye screening length). The grey triangle (b.c.c.) and squares (f.c.c.) are data for $1\text{-}\mu\text{m}$ spheres in cycloheptyl bromide-*cis*-decalin; (pure and at $c = 260 \mu\text{M}$); the circles are for $2\text{-}\mu\text{m}$ spheres in cyclohexyl bromide-*cis*-decalin, pure and $c = 10$ – $260 \mu\text{M}$. Open circles are fluid and closed circles are solid close to fluid–solid coexistence. Thick grey line is our best estimate of the phase boundary. **b–d**, x - y confocal images of: **b**, hexagonal plane in r.h.c.p. crystal; **c**, (100) face of f.c.c. (left) crystal coexisting with a (111) oriented f.c.c. (right); drawn in white is a grain boundary. **d**, (110) face (not a hexagonal plane) of b.c.c. crystal. **e**, Crystal structures are quantified by the in-plane pair correlation function $g(r)$ (plotted against r/a where $a \neq 2R$ is the mean interparticle spacing) and direct measurement of interplane stacking and separations. 2D $g(r)$ of b.c.c. (110) face (solid black line is experiment, grey vertical lines are calculated). Inset, a model of the b.c.c. (110) face. Grey and black circles denote different planes.

corresponding to $\varphi = 0.42\text{--}0.45$ at crystallization, which is hard-sphere-like (Fig. 2b) and close to the system reported in previous studies^{4,14}. The Debye lengths were estimated from conductivity measurements, implying a degree of salt dissociation of $<1\%$; electrophoresis measurements gave an estimate of $\zeta = +100\text{ mV} \approx 4k_B T/e$ for the surface potential (here k_B is Boltzmann's constant, T is temperature and e is the charge on the electron). We independently quantified the softness by the position of the first peak of the two-dimensional (2D) pair correlation function $g(r)$ as function of added salt at different values of φ , close to the fluid–solid phase boundary (see Methods for details).

The disorder–order phase diagram (Fig. 2a) as a function of κR and φ is illustrated by 2D planes from an r.h.c.p., f.c.c. and b.c.c. crystal (Fig. 2b, c and d, respectively). Samples with a linear concentration gradient (see Methods section for details) were used to map out phase boundaries. We then prepared homogeneous samples at volume fractions close to the phase boundary for a more quantitative analysis. A transition from a soft to a hard-sphere regime was observed (and quantified—see, for example, the analysis for the b.c.c. crystal in Fig. 2e). At $\kappa R \approx 0.3$, we obtained a ‘solid’ with a lattice spacing of $\approx 27\text{ }\mu\text{m}$ (Fig. 1c). At $\kappa R \approx 2$ we observed a fluid-to b.c.c.-to-f.c.c. phase sequence with increasing volume fraction, while at $\kappa R \approx 13$ and 26 a hard-sphere-like fluid-to-r.h.c.p. phase sequence was observed. Figure 1d shows the effect of the second deviation from hard-sphere-like, namely anisotropy. The electric field (along z in Fig. 1d, inset) caused the spheres to chain into strings along z . The average sphere–sphere spacing, σ_{chain} , in the string (obtained from a 2D autocorrelation of each xz snapshot) decreased (Fig. 1d) with increasing field, saturating when $(\sigma_{\text{chain}} - 2R)/2 = 100\text{ nm}$, a distance which is close to our estimate of the inverse-Debye length in the hard-sphere-like suspensions ($1/\kappa \approx 76\text{ nm}$). Fig. 1e shows $g(r)$ for a 2D snapshot in a 3D fluid and in a string fluid. Whereas the strings (such as shown in Fig. 1d inset) had one-dimensional order along z , the 2D plane

perpendicular to z was fluid-like, although with a stronger repulsion than in a normal fluid: the peak was higher and further from the origin.

We explored (Fig. 3) the phase diagram of the dipolar soft ($\kappa R \approx 2$) spheres (DP-SS) in the volume fraction–electric field plane. At low and zero electric fields, we observed a fluid-b.c.c.-f.c.c. phase sequence (with the b.c.c. phase existing for $0.08 < \varphi < 0.15$). The existence of the b.c.c. phase at volume fractions as high as $\varphi \approx 0.15$, although also observed in an X-ray scattering study of the phase diagram in an aqueous system²⁵ (also at comparable κR), and even feasible theoretically²⁷, is not expected for $\zeta/(k_B T/e) = 4$. For $0.08\text{ V }\mu\text{m}^{-1} < E_{\text{rms}} < 0.35\text{ V }\mu\text{m}^{-1}$, where E_{rms} is the root-mean-square electric field, strings were observed with a distribution of string lengths and monotonic-decreasing sphere separation (Fig. 1d and inset). At $E_{\text{rms}} = 0.35\text{ V }\mu\text{m}^{-1}$, with strings spanning the electrodes (and $\sigma_{\text{chain}} - 2R)/2 = 100\text{ nm}$), the ‘string-fluid’ was the stable phase for $\varphi < 0.15$. The long-range repulsion stabilized individual strings. It appears over the duration of our experiment (up to several hours) to be an equilibrium phase, with, as expected, the dynamics of the strings (Fig. 3e) being significantly slower (diffusion coefficient $D \approx 0.45(\mu\text{m})^2\text{ min}^{-1}$) than that of the individual spheres in the corresponding zero-field fluid phase. This is in sharp contrast with the dipolar hard-sphere (DP-HS) case, where strings form owing to the dipolar interaction but ‘zip up’ into sheets within a few milliseconds²⁰. The sheets

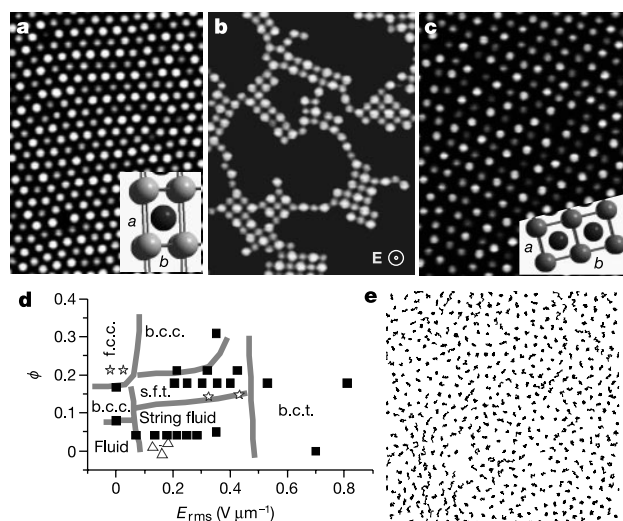


Figure 3 Volume fraction–electric field phase diagram. Data are for $1\text{-}\mu\text{m}$ spheres in cycloheptyl bromide-*cis*-decalin, no salt added. **a–c**, x - y confocal images of **a**, body-centred orthorhombic (b.o.) ($c = 2.187\text{ }\mu\text{m}$; $b = 1.244c$; $a = 1.97c$, model in inset), **b**, body-centred tetragonal (b.c.t.) ($a = b = 2.44\text{ }\mu\text{m}$; $c = 2.0\text{ }\mu\text{m}$, $a/c = 1.22$: ratio a/c for b.c.t. is theoretically $\sqrt{3}/2 = 1.2247$), and **c**, space-filling tetragonal (s.f.t.) crystal ($c = 2.0\text{ }\mu\text{m}$; $a = b$; $a/c = 1.846$, model in inset). **d**, The phase diagram. Grey lines indicate approximate phase boundaries, open stars two-phase coexistence, and open triangles are in the string fluid with a distribution of lengths of string that do not span the electrodes. The solid squares are all points where xyt or xyz data stacks were taken. **e**, Trajectory (time step, 30 s) of strings (ordered along z) diffusing in the x - y plane.

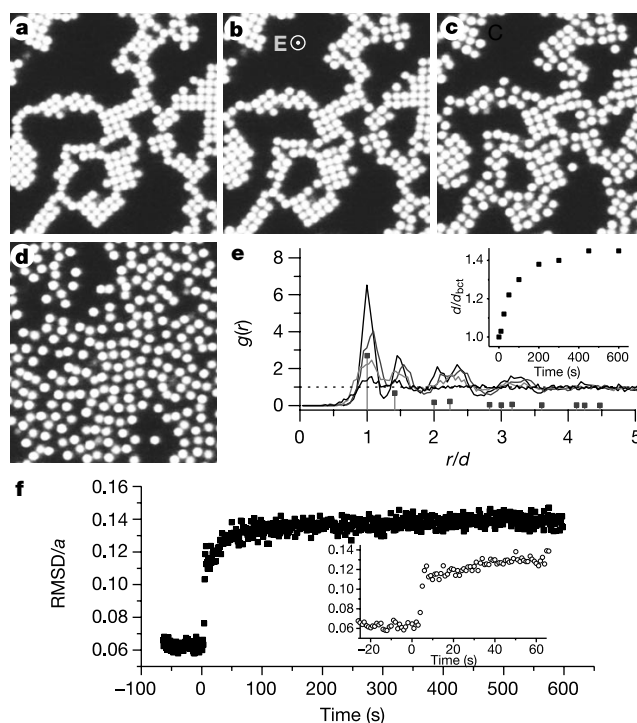


Figure 4 Melting of a b.c.t. crystal into a string fluid is not the reverse of crystallization of the string fluid. This quasi-2D transition is shown qualitatively in images taken at times $t = 0$ (**a**), 10 s (**b**), 50 s (**c**) and 600 s (**d**). The polycrystalline b.c.t. phase melts into a tetragonal crystal with larger in-plane lattice spacings: this crystal is metastable for several seconds. **e**, The $g(r)$ of two projected planes (which contain all the information in a unit cell) shows peaks characteristic of the initial four-fold symmetric lattice (grey vertical lines are calculated positions for a square lattice). Pair correlations at later times fall within the envelope of that at $t = 0\text{ s}$ provided that the distances are rescaled by a numerical factor that increases in time until it reaches the value for the peak position in the string fluid (inset). **f**, The scaled root-mean-square deviation (RMSD/ a) versus time; RMSD/ a is scaled with respect to the in-plane lattice constant ($a = b$) because the out-of-plane fluctuations are negligible. Within $t = 10\text{ s}$ (see inset), RMSD/ a increases from 0.06 to 0.12 , and at long times saturates at 0.14 .

too are short-lived dynamical structures that coarsen into three-dimensional b.c.t. ($a = b = c\sqrt{3}/2$) crystallites. We reproduced this DP-HS limit in a test sample of 1- μm spheres in cycloheptyl bromide/*cis*-decalin at $c = 260\ \mu\text{M}$, where we indeed observed a direct transition from fluid to (non-space-filling, $a = b = c\sqrt{3}/2$) b.c.t. crystallites with no intervening phase at intermediate fields.

Peculiar to DP-SS is an intermediate field regime with qualitatively different physics, and new crystal structures that arise purely from the competition of the soft repulsion and the anisotropic dipolar interaction. At $\phi \approx 0.25$ the string fluid crystallized to form a space-filling tetragonal (s.f.t.) solid phase (Fig. 3c), where (unlike the b.c.t. phase) the $a = b$ lattice vectors are such as to fill space, with $c \approx 2R$. Hence we observed, not several crystallites with the solvent continuum between them, but a single polycrystalline tetragonal solid. In order to distinguish the two, we use the terms s.f.t. and b.c.t. although in principle both structures are body-centred tetragonal. At $\phi > 0.30$, we observed (Fig. 3a) a body-centred orthogonal (b.c.o.) crystal structure $c \approx 2R \neq a \neq b$. Indeed there was a reversible transition from b.c.o. to s.f.t. as the field was increased at ($\phi = 0.21$, $E_{\text{rms}} = 0.32\ \text{V}\ \mu\text{m}^{-1}$). At high electric fields ($E_{\text{rms}} = 0.7\ \text{V}\ \mu\text{m}^{-1}$), we observed (Fig. 3b) b.c.t. crystallites with $a = b = c\sqrt{3}/2$, with $c \approx 2R$ as in the DP-HS case²⁰.

We can understand the physical processes at intermediate fields as follows. At low fields, charge repulsion dominates, whereas at high field, the anisotropic dipolar interaction (which is attractive along the field direction) predominates. At intermediate fields, attractions dominate along the field direction, while repulsion dominates in the plane perpendicular to this direction. Even in the DP-HS case, the attraction of two strings is only favoured when thermal fluctuations allow string undulations: in the absence of thermal fluctuations, the string-string interaction would decay exponentially²⁸. In the DP-SS case at intermediate fields, charge repulsion in the lateral plane dominates the weaker fluctuation effects.

Finally, we used the electric field as a surrogate temperature field to follow the dynamics of melting of a b.c.t. crystal into a string-fluid phase close to the string-fluid-s.f.t. phase boundary ($\phi = 0.1$). On increasing the field, the string fluid crystallized into b.c.t. ($a = b = 2R\sqrt{3}/2$). On lowering the field, we saw the sequence shown in Fig. 4. Figure 4a shows the b.c.t. when the field was reduced from $E_{\text{rms}} = 0.53\ \text{V}\ \mu\text{m}^{-1}$ to $0.35\ \text{V}\ \mu\text{m}^{-1}$. At $t = 10\ \text{s}$ and $25\ \text{s}$, there was still crystallinity with four-fold symmetry (as shown in Fig. 4e where the $g(r)$ is plotted for all the times shown in the images, and compared with the peaks expected for a perfect square lattice), which decreased in time. Finally at $t = 600\ \text{s}$, the structure was almost liquid-like. Note that in Fig. 4e, the distance is rescaled such that the first peak is unity. This rescaling factor (see inset) was $d = d_{\text{bct}} = (\sqrt{3}/2)(2R)$ for a projection of two consecutive (001) b.c.t. planes, but increased with time until it saturates at the string-fluid value. The four-fold symmetric $g(r)$ persisted with this larger transverse length scale for more than $25\ \text{s}$. We also computed a dynamical order parameter, the average root-mean-square deviation (RMSD/ a , scaled with respect to the in-plane lattice constant; Fig. 4f) of particle positions between two successive images in the time series. Here, within $t = 10\ \text{s}$, RMSD/ a increased from 0.06 to 0.12, and at long times saturated at 0.14. Thus the b.c.t. phase melted quickly ($t < 10\ \text{s}$), but the equilibrium string-fluid phase was preceded by a transient s.f.t. structure that persisted for minutes.

The present system significantly extends the use of colloids as condensed matter model systems. The system is refractive-index and density matched, so that bulk measurements in real-space are possible. In the intermediate-field regime of the phase diagram where both softness and anisotropy are important, we observed three colloidal phases composed of repelling strings. Long-ranged repulsive spheres are being used in new photonic applications²⁹, and dipolar interactions in electro-rheological fluids. We expect that the phases we report here, together with the increased control over

interparticle interactions, that is now possible, will lead to important extensions of these and other advanced materials. □

Methods

Sample preparation

Special precautions need to be taken when using low-ionic-strength non-aqueous suspensions. The cyclohexyl bromide ('CH6B'), cycloheptyl bromide ('CH7B') and *cis*-decalin ('CD') were obtained from Sigma-Aldrich. The cycloheptyl and hexyl bromides had a yellowish colour (as bought), and were dried first and then vacuum-distilled, resulting in a colourless liquid. The *cis*-decalin was used as bought. The solutions of tetrabutylammonium chloride ('TBAC'; Sigma-Aldrich) in CH6B or CH7B were prepared by dilution from a stock solution which was close to the saturation concentration at room temperature (293 K) and pressure.

Transient charging effects caused by tribo-electricity can have relaxation times of the order of minutes, so particle suspensions were allowed to stand for hours to be sure they were in equilibrium. Moreover, we could eliminate static charging effects by using indium tin oxide (ITO)-coated electrodes, and shorting and grounding both electrodes to be sure that the electric field was truly zero. In the experiments reported here, we imaged primarily at low intensities (the 488 nm line of the Ar/Kr mixed gas laser at $< 200\ \mu\text{W}$) to make sure the self-dissociation of the liquid was not influenced by the light. We used both rectangular glass capillaries and sandwich ITO glass cells.

Electrical conductivity measurements

We measured the electric conductivity (WTW InoLab Level 1, with LR325/01 conductivity cell; cell constant $0.100\ \text{cm}^{-1}$) of pure CH6B-CD (and CH7B-CD) and for salt concentrations $c = 10$ – $260\ \mu\text{M}$. Although the conductivity in the pure solvent mixture at all particle concentrations was too low ($< 0.02\ \mu\text{S}\ \text{cm}^{-1}$) to measure, we could measure conductivities with salt added. For these values we may then estimate the ionic strength c_i using Walden's rule ($\Lambda_0^{\text{water}} = \Lambda_0 \eta_0$, where Λ_0 , η_0 , and Λ_0^{water} , η_0^{water} are the limiting molar conductivity and solvent viscosity in the solvent of interest and in water, respectively), $c_i = K/(10^3 \Lambda_0)$, K being the electrical conductivity, and thereby the Debye length $\kappa^{-1} = 0.034\sqrt{\epsilon/c_i}$. For example, at salt concentration $c = 260\ \mu\text{M}$ (by weighing) we obtained an ionic strength $c_i = 1.2\ \mu\text{M}$ giving an inverse Debye length of 76 nm. The degree of salt dissociation is thus $c_i/c \approx 0.5\%$.

Electrophoretic mobility measurements

To further quantify the softness, we measured the electrophoretic mobility via electrophoretic light scattering (Coulter Delsa 440SX) in a density-matched but highly scattering solvent mixture (CH6B-*n*-hexane), which exhibits similar charged behaviour: owing to the low electrical conductivity in the zero-salt solvent mixture, this was done only at $c_i \approx 0.2\ \mu\text{M}$. In this case we obtained a mobility of $0.37\ \mu\text{m}\ \text{cm}\ \text{V}^{-1}\ \text{s}^{-1}$ which yields (via the O'Brien and White numerical scheme³⁰) $\zeta \approx 4k_B T/e$, a rather large and positive surface potential. Given that the inverse Debye length obtained is a lower bound, the ζ value is a lower bound as well.

Phase boundary determination

Phase boundaries were first mapped out using concentration gradient samples. Such samples were made by filling half of a 100- μm -thick capillary with suspension at large ϕ , then the other half at low ϕ , sealing, allowing to equilibrate for several days, and looking closely at the interfacial region. Another way of accessing concentration gradients very quickly was to look at the sedimentation profile in imperfectly density-matched suspensions. The interparticle spacing at the fluid-solid interface gave us a rough but dependable estimate of ϕ . Via confocal microscopy we obtained 2D slices in the sample bulk, which we used to calculate the 2D pair correlation function $g(r)$ (for example, Figs 1e and 2e). The position of the first peak of the pair-correlation function was an independent determinant of the crystallization volume fraction at the fluid-solid phase boundary, and therefore an independent quantifier of the interaction softness.

Received 15 August; accepted 19 November 2002; doi:10.1038/nature01328.

- Grier, D. G. & Murray, C. A. The microscopic dynamics of freezing in supercooled colloidal fluids. *J. Chem. Phys.* **100**, 9088–9095 (1994).
- van Blaaderen, A. & Wiltzius, P. Real-space structure of colloidal hard-sphere glasses. *Science* **270**, 1177–1179 (1995).
- Kegel, W. K. & van Blaaderen, A. Direct observation of dynamical heterogeneities in colloidal hard-sphere suspensions. *Science* **287**, 290–293 (2000).
- Gasser, U., Weeks, E. R., Schofield, A., Pusey, P. N. & Weitz, D. A. Real-space imaging of nucleation and growth in colloidal crystallization. *Science* **292**, 258–262 (2001).
- Antl, L., Goodwin, J., Hill, R., Ottewill, R. & Waters, J. The preparation of poly(methyl methacrylate) latices in non-aqueous media. *Colloids Surf.* **17**, 67–78 (1986).
- van Blaaderen, A. & Vrij, A. Synthesis and characterization of colloidal dispersions of fluorescent, monodisperse silica spheres. *Langmuir* **8**, 2921–2931 (1992).
- Dhont, J. K. G., Smits, C. & Lekkerkerker, H. N. W. A time resolved static light-scattering study on nucleation and crystallization in a colloidal system. *J. Colloid Interf. Sci.* **152**, 386–401 (1992).
- Russel, W. B., Chaikin, P. M., Zhu, J., Meyer, W. V. & Rogers, R. Dendritic growth of hard sphere crystals. *Langmuir* **13**, 3871–3881 (1997).
- Palberg, T. Crystallization kinetics of repulsive colloidal spheres. *J. Phys. Condens. Matter* **11**, R323–R360 (1999).
- Harland, J. L. & vanMegen, W. Crystallization kinetics of suspensions of hard colloidal spheres. *Phys. Rev. E* **55**, 3054–3067 (1997).
- Gast, A. P. & Russel, W. B. Simple ordering in complex fluids—Colloidal particles suspended in solution provide intriguing models for studying phase transitions. *Phys. Today* **51**, 24–30 (1998).

12. Pham, K. N. *et al.* Multiple glassy states in a simple model system. *Science* **269**, 104–106 (2002).
13. Weeks, E. R., Crocker, J. C., Levitt, A. C., Schofield, A. & Weitz, D. A. Three-dimensional direct imaging of structural relaxation near the colloidal glass transition. *Science* **287**, 627–631 (2000).
14. Pusey, P. N. & van Meegen, W. Phase behaviour of concentrated suspensions of nearly hard colloidal spheres. *Nature* **320**, 340–342 (1986).
15. Russel, W. B., Schowalter, W. R. & Saville, D. A. *Colloidal Dispersions* (Cambridge Univ. Press, Cambridge, 1999).
16. Parthasarathy, M. & Klingenberg, D. J. Electrorheology: mechanisms and models. *Mater. Sci. Eng.* **R17**, 57–103 (1996).
17. Tao, R. & Jiang, Q. Simulation of structure formation in an electrorheological fluid. *Phys. Rev. Lett.* **73**, 205–208 (1994).
18. Martin, J. E., Odinek, J. & Halsey, T. C. Evolution of structure in a quiescent electrorheological fluid. *Phys. Rev. Lett.* **69**, 1524–1527 (1992).
19. Martin, J. E., Anderson, R. A. & Tigges, C. P. Simulation of the athermal coarsening of composites structured by a uniaxial field. *J. Chem. Phys.* **108**, 3765–3787 (1998).
20. Dassanayake, U., Fraden, S. & van Blaaderen, A. Structure of electrorheological fluids. *J. Chem. Phys.* **112**, 3851–3858 (2000).
21. Bosma, G. *et al.* Preparation of monodisperse, fluorescent PMMA-latex colloids by dispersion polymerization. *J. Colloid Interf. Sci.* **245**, 292–300 (2002).
22. de Hoog, E. H. A., Kegels, W. K., van Blaaderen, A. & Lekkerkerker, H. N. W. Direct observation of crystallization and aggregation in a phase-separating colloid-polymer suspension. *Phys. Rev. E* **64**, 021407 (2001).
23. Pronk, S. & Frenkel, D. Can stacking faults in hard-sphere crystals anneal out spontaneously? *J. Chem. Phys.* **110**, 4589–4592 (1999).
24. Robbins, M. O., Kremer, K. & Grest, G. S. Phase diagram and dynamics of Yukawa systems. *J. Chem. Phys.* **88**, 3286–3312 (1988).
25. Sirota, E. B. *et al.* Complete phase-diagram of a charged colloidal system - A synchrotron X-ray scattering study. *Phys. Rev. Lett.* **62**, 1524–1527 (1989).
26. Monovoukas, Y. & Gast, A. P. The experimental phase diagram of charged colloidal suspensions. *J. Colloid Interf. Sci.* **128**, 533–548 (1989).
27. El Azhar, F., Baus, M., Ryckaert, J.-P. & Meijer, E. J. Line of triple points for the hard-core Yukawa model: A computer simulation study. **112**, 5121–5126 (2000).
28. Halsey, T. C. & Toor, W. Structure of electrorheological fluids. *Phys. Rev. Lett.* **65**, 2820–2823 (1990).
29. Holtz, J. H. & Asher, S. A. Polymerized colloidal crystal hydrogel films as intelligent chemical sensing materials. *Nature* **389**, 829–832 (1997).
30. O'Brien, R. W. & White, L. R. Electrophoretic mobility of a spherical colloidal particle. *J. Chem. Soc. Faraday Trans. II* **74**, 1607–1626 (1978).

Acknowledgements We thank G. Bosma for particle synthesis, C. van Kats for electrophoresis measurements, H. Wisman for technical support, K. van Walree for the suggestion of TCAB salt, and discussion of charge mechanisms, and J. Hoogenboom, S. Auer and D. Frenkel for discussion. We also thank M. Dijkstra and G. Patey for a critical reading of the manuscript. This work is part of the research program of the 'Stichting voor Fundamenteel Onderzoek der Materie (FOM)', which is financially supported by the 'Nederlandse organisatie voor Wetenschappelijke Onderzoek (NWO)'.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.Y. (e-mail: yethiraj@chem.ubc.ca) or A.v.B. (e-mail: A.vanBlaaderen@phys.uu.nl).

High-temperature superconductor bulk magnets that can trap magnetic fields of over 17 tesla at 29 K

Masaru Tomita*† & Masato Murakami*

* Superconductivity Research Laboratory, Shibaura 1-16-25, Minato-ku, Tokyo 105-0023, Japan

† Railway Technical Research Institute, 2-8-38 Hikari-cho, Kokubunji-shi, Tokyo 185-8540, Japan

Large-grain high-temperature superconductors of the form RE-Ba-Cu-O (where RE is a rare-earth element) can trap magnetic fields of several tesla at low temperatures, and so can be used for permanent magnet applications^{1,2}. The magnitude of the trapped field is proportional to the critical current density and the volume of the superconductor^{3,4}. Various potential engineering applications for such magnets have emerged^{5–13}, and some have already been commercialized^{7–10}. However, the range of applications is limited by poor mechanical stability and low

thermal conductivity of the bulk superconductors^{14–17}; RE-Ba-Cu-O magnets have been found to fracture during high-field activation, owing to magnetic pressure^{14–16}. Here we present a post-fabrication treatment that improves the mechanical properties as well as thermal conductivity of a bulk Y-Ba-Cu-O magnet, thereby increasing its field-trapping capacity. First, resin impregnation and wrapping the materials in carbon fibre improves the mechanical properties. Second, a small hole drilled into the centre of the magnet allows impregnation of Bi-Pb-Sn-Cd alloy into the superconductor and inclusion of an aluminium wire support, which results in a significant enhancement of thermal stability and internal mechanical strength. As a result, 17.24 T could be trapped, without fracturing, in a bulk Y-Ba-Cu-O sample of 2.65 cm diameter at 29 K.

Bulk RE-Ba-Cu-O superconductors have a complex structure, consisting of an REBa₂Cu₃O₇ matrix in which are distributed small particles of RE₂BaCuO₅. REBa₂Cu₃O₇ materials undergo the tetragonal to orthorhombic transformation, which causes crystal deformation. During this transformation the crystal takes up oxygen, and as a result the *c* axis shrinks, leading to the formation of microcracks perpendicular to this axis¹⁸. In addition, when the RE-Ba-Cu-O precursors melt, oxygen gas is released from the crystal; some of this gas remains in the sample (because of high viscosity), resulting in the formation of microvoids¹⁹. These defects are difficult to elim-

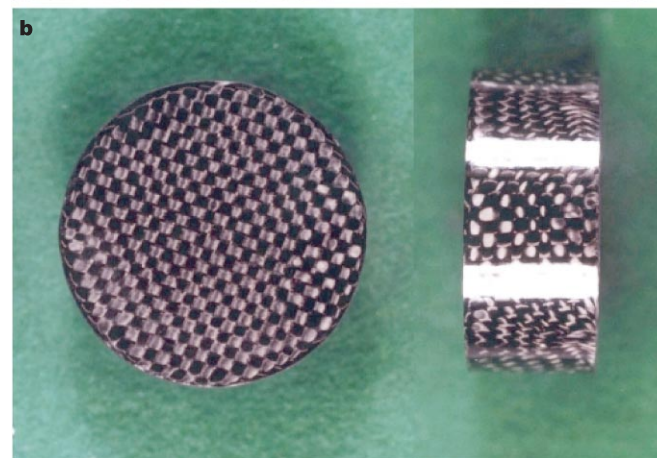
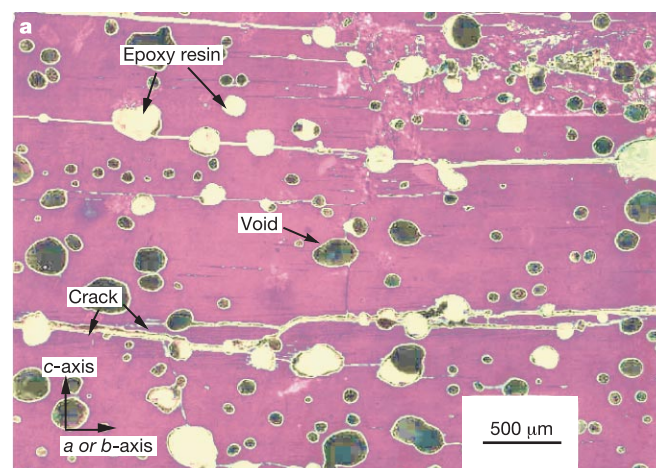


Figure 1 Resin-impregnated Y-Ba-Cu-O with carbon fibre wrapping. **a**, Cross-sectional view of a resin-impregnated Y-Ba-Cu-O (YBCO) disk. Note that resin penetrates through the surface cracks, and fills cracks and voids inside the sample. **b**, A photograph of a YBCO disk wrapped with carbon fibre fabric and resin-impregnated. The stripes of carbon fibres are clearly visible. The carbon fibre reduced the thermal expansion coefficient, and further enhanced the mechanical strength of the resin.

inate, and cause a deterioration of the mechanical properties. Thus the tensile strength of bulk Y-Ba-Cu-O (YBCO) along the *a*-*b* plane is relatively low, ranging from 10 to 30 MPa (ref. 20).

The interaction of the trapped field and the current causes an outward pressure proportional to the trapped field. Therefore, the tensile strength (σ_B) sets the maximum trapped field ($B_{\max}$), which is to first approximation given by the following equation²¹:

$$\sigma_B = 0.282 B_{\max}^2 \quad (1)$$

where σ_B is in MPa, and $B_{\max}$ is in T. According to this relation, the maximum trapped field ranges from 6.0 to 9.4 T in bare bulk superconductors.

Addition of silver has been found effective in improving the mechanical properties, producing a slight increase in the maximum trapped field²². Reinforcement of the sample with metal rings has also been found to be effective in improving the mechanical properties, because the bulk is compressed owing to a difference in thermal contraction^{16,23}. In fact, a trapped field of 14.35 T has been achieved at 22.5 K, although the sample fractured after experiments¹⁶. However, these techniques could not improve the internal mechanical strength or cryo-stability of bulk superconductors.

We have recently found²⁴ that resin can penetrate into a bulk superconductor. When the present sample of YBCO was immersed in molten resin, the resin permeated into the bulk through micro-cracks having openings on the surface (Fig. 1a). The voids connected to these cracks were also filled with resin, which dramatically enhanced the tensile strength from 18.4 to 77.4 MPa (ref. 25). However, a large difference in the thermal expansion coefficient between YBCO ($1 \times 10^{-5} \text{ K}^{-1}$) and resin ($3\text{--}4.5 \times 10^{-5} \text{ K}^{-1}$) caused damage in the surface resin layer during thermal cycles. We therefore wrapped the YBCO disk with carbon fibre fabric before resin impregnation, as carbon fibre has a smaller thermal expansion coefficient of $2 \times 10^{-5} \text{ K}^{-1}$, and can avoid the excessive thermal contraction of the resin.

We measured the trapped field between two resin-impregnated YBCO disks 2.65 cm in diameter and 1.5 cm in thickness, fabricated with the top-seeded melt-growth process²⁶. Each disk was initially characterized by measuring the trapped field profile at 77 K by

cooling the sample with liquid nitrogen in an applied field of 1.5 T (using an normal conducting electromagnet). After removal of the applied field, an axial Hall sensor (model BHA 921, Bell) was scanned over the entire surface, at a constant sensor-surface distance of 1.2 mm. The two disks showed a maximum field of about 0.7 T with a single peak. The YBCO disks were wrapped with carbon fibre fabric, and placed in a plastic mould. Epoxy resin was melted at 70 °C, admitted to the mould, and held for 1 h; outgassing gases were evacuated with a rotary pump. The epoxy resin was composed of phenol-formaldehyde, with polyamine added as hardening material in a weight ratio of 100:32. Finally the mould assembly was heated at 80 °C for 6 h, and at 120 °C for 2 h in air to cure the resin. Figure 1b shows a photograph of a YBCO disk wrapped with carbon fibre fabric and impregnated with resin. The mesh structure of carbon fibre fabrics is visible through the surface resin. No damage occurred in the resin during thermal cycles, thanks to carbon fibre treatment.

For trapped-field measurements, we used an 18-T superconducting hybrid (NbTi and Nb₃Sn) magnet with a bore diameter of 3.6 cm at the Tsukuba Magnet Laboratory of the National Research Institute for Materials. An axial cryogenic Hall sensor (model LHP-MU, AREPOC) was sandwiched by two YBCO disks at their centre to monitor the field. The two disks with the Hall sensor were glued together with GE varnish, and further glued to the cold stage of a variable-temperature cryostat, which allowed us to vary the temperature between 4.2 and 300 K with helium gas circulation. Between two disks, a temperature sensor (model CX-1070 SD, Lake Shore Cernox) was also inserted to monitor the temperature. For each test, the sample was first held at 100 K (that is, above the superconducting transition temperature of YBCO) while the superconducting magnet was energized to a maximum field of 17.9 T. The sample was then cooled to 29 K. The field and temperature of the samples were monitored as the applied field was swept down to zero at a rate of 0.15 T min^{-1} .

Figure 2 shows the measured magnetic field at the centre of the two Y-Ba-Cu-O disks while the field was decreased from 17.9 T to zero. The temperature change of the bulk surface is also shown. Although the temperature of the cryostat was set at 29 K, the temperature of the sample did not reach this value and was 31 K, owing to low thermal conductivity. As the applied field was swept down, the temperature of the sample immediately rose by two degrees and kept increasing. To avoid flux avalanche, we lowered the sweep rate to 0.1 T min^{-1} when the applied field reached 10 T. The field stayed at 17.9 T at the centre of the sample until the applied field reached 6 T, reflecting an extremely strong flux pinning force. Despite careful control of the field sweeping rate, the temperature of the sample suddenly rose to $>70 \text{ K}$ at an applied field of 6.0 T and

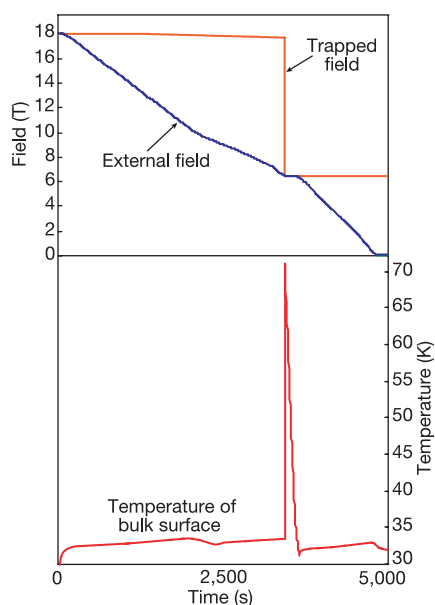


Figure 2 Trapped-field data for two YBCO disks with carbon fibre wrapping and resin impregnation. Shown are the magnetic field between the two disks, the external field, and the temperature of the YBCO disk magnet, as a function of time, when the applied field is decreased from 17.9 T to zero.

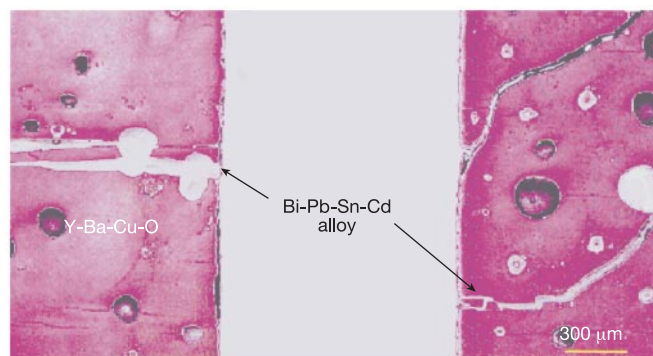


Figure 3 Cross-sectional view of a YBCO disk in which a hole of 1 mm diameter was drilled, followed by Bi-Pb-Sn-Cd alloy impregnation. Note that voids and cracks, along with the hole, are filled with alloy.

the field strength on the bulk surface also dropped from 17.9 to 6.0 T. We then stopped sweeping the field, and monitored the sample temperature. After three minutes the temperature of the sample recovered to 31 K. As the field was decreased further, no temperature spike occurred and the field strength of the bulk surface stayed at 6 T even when the external field reached zero.

After the experiment, the sample was subjected to visual inspection. A large crack was observed on both YBCO disks. We note that a sudden change in field distribution in YBCO disks can be explained in terms of flux jumping; such thermal instability is commonly observed in low-temperature superconductors²⁷. The problem here was that the heat generated in the superconductor could not be effectively transferred to the cooling device. We then performed further similar experiments, but with the sweep rate reduced to $<0.05 \text{ T min}^{-1}$ in a manual mode to allow longer time for heat transfer. We could achieve a trapped field of 14.9 T at 31 K, but the trapped field magnet was unstable and small agitation caused flux jumping. As a result, all the samples were destroyed after high field activation. Similar problems of sample cracking were encountered in previous experiments^{9,14–16}.

The thermal conductivity of bulk YBCO is anisotropic: 3.5 W m K^{-1} along the *c* axis, and 14 W m K^{-1} along the *a–b* plane. We believe that such a low thermal conductivity causes flux jumping. Thus the key to cryo-stability is heat transfer to the cooling device.

In order to overcome this problem, we impregnated the bulk sample with $\text{Bi}_{0.5}\text{-Pb}_{0.27}\text{-Sn}_{0.13}\text{-Cd}_{0.1}$ (Bi-Pb-Sn-Cd) alloy, which has a high thermal conductivity of 200 W m K^{-1} at 29 K. This was achieved with the same method as used for epoxy resin impreg-

nation, as the melting point of this alloy is only 70°C . Another reason for the selection of this alloy was that its thermal expansion coefficient ($2 \times 10^{-5} \text{ K}^{-1}$) is close to that of bulk YBCO, so that thermal stresses could be made negligibly small. This treatment markedly enhanced the thermal conductivity, but flux jump was still unavoidable. This is because the alloy impregnation could enhance the thermal conductivity only at the surface, while the thermal conductivity of the interior region of the disk remained low.

We sought a technique to improve the thermal conductivity inside the sample. To achieve this, we mechanically drilled a hole 1 mm in diameter at the centre of the sample, and followed this by Bi-Pb-Sn-Cd alloy impregnation. This mechanical machining did not degrade the performance of the YBCO disk. Figure 3 is an optical micrograph of the cross-section along the hole for the alloy-impregnated sample. We note that microcracks as well as voids can be filled with Bi-Pb-Sn-Cd alloy through the hole drilled in the sample.

With the aim of full stabilization, we prepared YBCO disk magnets with the following steps. First, two disks were subjected to carbon fibre/resin impregnation with the top and bottom surfaces uncovered. A hole of 1 mm diameter was drilled at the centre of each disk. Then we inserted 0.9-mm-diameter Al wire, of large thermal conductivity ($4,000 \text{ W m K}^{-1}$ at 29 K), into the hole for further enhancement of thermal conductivity. The disks were then placed in a carbon tube, leaving a minimum space around the disk. Bi-Pb-Sn-Cd was melted at 70°C and admitted to the mould.

This configuration was chosen because the surfaces free from the coolant should be thermally insulated to prevent heat leaks. Two YBCO disks were stacked together with GE varnish. This time, six Hall sensors (model LHP-MU, AREPOC) were sandwiched by two YBCO disks to monitor the field distribution at the following locations: the centre, 0.06 cm (two locations), 0.68 cm, 1.01 cm and 1.32 cm away from the centre. This assembly was subjected to trapped field measurements at the sweeping rate of 0.15 T min^{-1} .

Figure 4 shows variation of the field and temperature in the present composite bulk magnet when the applied field was swept

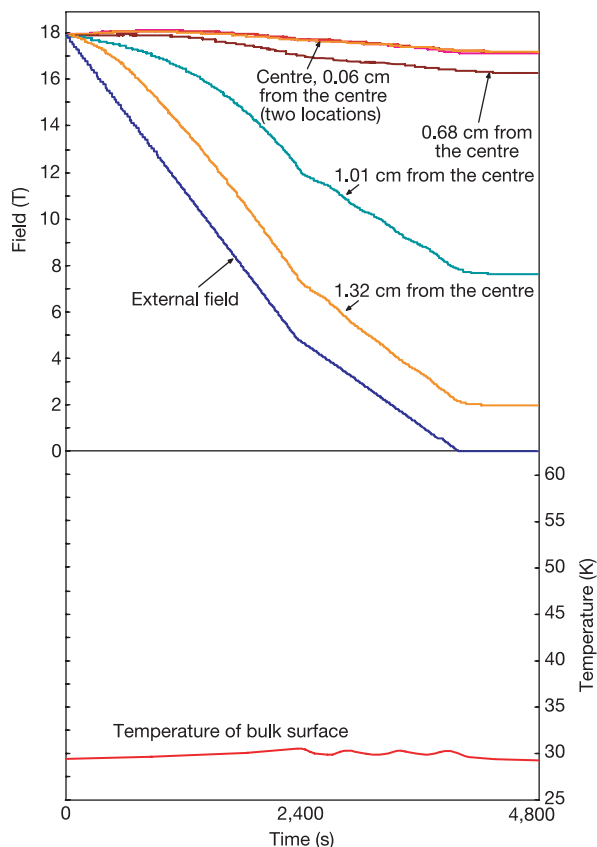


Figure 4 Trapped-field data for two fibre-wrapped, resin-impregnated YBCO disks with embedded Al wire. Shown are the variation of the trapped-field, the external field, and the temperature, for the bulk disks when the external field is reduced from 17.9 T to zero.

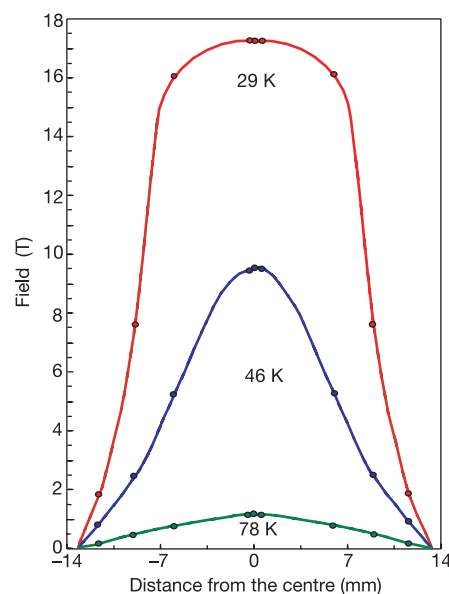


Figure 5 The effect of temperature on trapped-field distribution. The field was trapped between two 26.5-mm-diameter YBCO disks with carbon fibre wrapping, resin impregnation and embedded Al. Data are shown for 29 K, 46 K and 78 K. It is evident that the trapped field is saturated at higher temperatures, but that the field is far below saturation at 29 K, showing that much higher fields could be trapped.

down from 17.90 T to zero. The temperature of the sample reached 29 K, reflecting a marked enhancement of its thermal conductivity. Also, the temperature rise during field sweeping is small compared to the experiment shown in Fig. 2. This also supports the fact that the thermal conductivity was greatly improved in the present configuration. As the applied field is decreased, the field on the superconductor gradually decreased. However, the field at the centre of the sample stayed at an initial value of 17.9 T. No sudden temperature rise occurred during the field-decreasing process, but, because the temperature increased to 30 K at 4 T, we lowered the sweeping rate to provide a safety margin. Even after the external field was removed, a field of >17 T was trapped by the superconductor (Fig. 5). We also performed similar trapped-field experiments at 78 K and 46 K (Fig. 5). The trapped-field profiles at these temperatures show saturation, with a peak field of 9.5 T for 46 K and 1.2 T for 78 K. Compared to the data at 46 and 78 K, the trapped field at 29 K is far below saturation, and hence the present bulk magnet has the potential to trap much higher fields than 17 T if a higher magnetic field is available for activation.

The static field of 17.24 T is attractive for various applications. As long as the sample temperature is kept at 29 K with a cryocooler, a bulk superconductor magnet trapping 17.24 T could be transported to any desired place. This means that a strong static field of 17.24 T limited to the bore space of the superconducting solenoid could be freely used in free space. For magnetic separation of contaminants from polluted water, pre-treatment with ferromagnetic powders is unnecessary for fields larger than 10 T. Therefore, a much simpler separation system could be made. Furthermore, the output of an electromagnetic machine is the product of electric current and magnetic field, so a strong magnetic field is beneficial for the development of high-power, compact machines. The range of a stray field is very small in bulk superconductor magnets, so the deleterious effect of stray field can be localized. We therefore believe that the present results will open the way to new applications of fields from bulk superconducting magnets. □

Received 21 August; accepted 29 November 2002; doi:10.1038/nature01350.

- Nariki, S., Sakai, N. & Murakami, M. Processing of high-performance Gd-Ba-Cu-O bulk superconductor with Ag addition. *Supercond. Sci. Technol.* **15**, 648–652 (2002).
- Ikuta, H. *et al.* Melt-processed Sm-Ba-Cu-O superconductors trapping strong magnetic field. *Supercond. Sci. Technol.* **11**, 1345–1347 (1998).
- Murakami, M. *Melt Processed High Temperature Superconductors* (World Scientific, Singapore, 1991).
- Cardwell, D. A. Processing and properties of large grain (RE)BCO. *Mater. Sci. Eng. B* **53**, 1–10 (1998).
- Weinberger, B. R., Lynds, L., Hull, J. R. & Balachandran, U. Low friction in high temperature superconductor bearings. *Appl. Phys. Lett.* **59**, 1132–1134 (1991).
- Matsunaga, M. *et al.* YBCO bulk for the superconducting bearing for a 10kWh flywheel. *Supercond. Sci. Technol.* **15**, 842–845 (2002).
- Minami, H. & Yuyama, J. Construction and performance test of a magnetically levitated transport system in vacuum using high- T_c superconductors. *Jpn J. Appl. Phys.* **34**, 346–349 (1995).
- Kovalev, L. K. *et al.* High output power reluctance electric motors with bulk high-temperature superconductor elements. *Supercond. Sci. Technol.* **15**, 817–822 (2002).
- Oka, T. *et al.* Construction of 2–5 T class superconducting magnetic field generator with use of an Sm123 bulk superconductor and its application to high-magnetic field demanding devices. *Physica C* **335**, 101–106 (2000).
- Saho, N., Mizumori, T., Nishijima, N., Murakami, M. & Tomita, M. High- T_c bulk-superconductor based membrane-magnetic separation for water purification. *J. Am. Ceram. Soc.* (in the press).
- Mizutani, U. *et al.* Application of RE123-bulk superconductors as a permanent magnet in magnetron sputtering film deposition apparatus. *J. Am. Ceram. Soc.* (in the press).
- Yoo, S. I., Higuchi, T., Sakai, N., Fujimoto, H. & Murakami, M. RE-Ba-Cu-O for high functional superconducting permanent magnet. *Mater. Sci. Eng. B* **52**, 203–210 (1998).
- Wang, J. *et al.* A scheme of maglev vehicle using high T_c bulk superconductors. *IEEE Trans. Appl. Supercond.* **9**, 904–907 (1999).
- Ren, Y., Weinstein, R., Liu, J., Sawh, R. P. & Foster, C. Damage caused by magnetic pressure at high trapped field in quasi-permanent magnets composed of melt-textured Y-Ba-Cu-O superconductor. *Physica C* **251**, 15–26 (1995).
- Gao, L., Xue, Y., Meng, L. & Chu, C. W. Thermal instability, magnetic shielding and trapping in single-grain YBa₂Cu₃O_{7- δ} bulk materials. *Appl. Phys. Lett.* **64**, 520–522 (1994).
- Fuchs, G. *et al.* Trapped magnetic fields larger than 14 T in bulk YBa₂Cu₃O_{7- δ} . *Appl. Phys. Lett.* **76**, 2107–2109 (2000).
- Johansen, T. H. Flux-pinning-induced stress and strain in superconductors: Case of a long circular cylinder. *Phys. Rev. B* **60**, 9690–9713 (1999).
- Diko, P. Thermal stresses and microcracking caused by 211 particles in Y-Ba-Cu-O melt-processed bulks. *Mater. Sci. Eng. B* **53**, 149–153 (1998).

- Sakai, N., Ishihara, D., Inoue, I. & Murakami, M. Formation of pores in melt-processed RE-Ba-Cu-O and the techniques to reduce pore density. *Supercond. Sci. Technol.* **15**, 698–701 (2002).
- Sakai, N., Seo, S. J., Inoue, K., Miyamoto, T. & Murakami, M. in *Advances in Superconductivity XI* (eds Koshizuka, N. & Tajima, S.) 685–688 (Springer, Tokyo, 1999).
- Collings, E. W. & Sumption, M. D. in *Advances in Superconductivity VII* (eds Yamafuji, K. & Morishita, T.) 605–608 (Springer, Tokyo, 1995).
- Yeh, F. & White, K. W. Fracture toughness behavior of the YBa₂Cu₃O_{7- δ} superconducting ceramic with silver oxide additions. *J. Appl. Phys.* **70**, 4989–4994 (1991).
- Morita, M. *et al.* Trapped field and mechanical properties of QMG magnet. *Proc. 1988 Int. Workshop on Superconductivity* 115–118 (ISTEC, Tokyo, 1998).
- Tomita, M. & Murakami, M. Improvement of the mechanical properties of bulk superconductors with resin impregnation. *Supercond. Sci. Technol.* **13**, 722–724 (2000).
- Tomita, M., Murakami, M. & Katagiri, K. Reliability of mechanical properties for bulk superconductors with resin impregnation. *Physica C* **378–381**, 783–787 (2002).
- Salama, K. *et al.* Recent developments in melt-textured superconductors. *JOM* **17–28** (June 2000).
- Iwasa, Y. *Case Studies in Superconducting Magnets* (Plenum, New York, 1994).

Acknowledgements We thank K. Itoh and H. Wada for use of an 18-T superconducting magnet. This work was partially supported by NEDO as Collaborative Research and Development of Fundamental Technologies for Superconductivity Applications.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.T. (e-mail: tomita@rtri.or.jp)

The effects of surfactants on spilling breaking waves

Xinan Liu & James H. Duncan

Department of Mechanical Engineering, University of Maryland, College Park, Maryland 20742, USA

Breaking waves markedly increase the rates of air–sea transfer of momentum, energy and mass^{1–4}. In light to moderate wind conditions, spilling breakers with short wavelengths are observed frequently. Theory and laboratory experiments have shown that, as these waves approach breaking in clean water, a ripple pattern that is dominated by surface tension forms at the crest^{5–14}. Under laboratory conditions and in theory, the transition to turbulent flow is triggered by flow separation under the ripples, typically without leading to overturning of the free surface¹⁵. Water surfaces in nature, however, are typically contaminated by surfactant films that alter the surface tension and produce surface elasticity and viscosity^{16,17}. Here we present the results of laboratory experiments in which spilling breaking waves were generated mechanically in water with a range of surfactant concentrations. We find significant changes in the breaking process owing to surfactants. At the highest concentration of surfactants, a small plunging jet issues from the front face of the wave at a point below the wave crest and entraps a pocket of air on impact with the front face of the wave. The bubbles and turbulence created during this process are likely to increase air–sea transfer.

Experiments were done in a tank of 15 m by 1.2 m with a water depth of 0.8 m. Weak spilling breakers were produced from Froude-scaled mechanically generated dispersively focused wave packets^{18,19} with average frequencies of 1.15, 1.26 and 1.42 Hz. The breaker generation technique and wave-maker motion parameters are identical to those described in ref. 13 (in the present case, the dimensionless wave-maker amplitude is 0.0505). The wave-maker motions are repeatable to within $\pm 0.1\%$ of their amplitude.

All of the experiments reported here were done with a single tank of filtered tap water over a period of 8 d. Rhodamine 6G, a fluorescent dye, was mixed with the water at a concentration of

0.5 p.p.m. for visualization purposes. The main surfactant used in this work was the soluble surfactant sodium dodecyl sulphate (SDS), although other surfactants (from bacteria, impurities in the chemicals and other sources) would have been present as well. Measurements were done with relatively clean water (data set CLEAN) and five different concentrations of surfactants (data sets 1, 2, 3, 4a and 4b). The bulk concentration of SDS (C) corresponding to each data set is given in Fig. 1. The SDS concentrations in data sets 4a and 4b are equal, but the measurements were taken 3 d apart (see Methods). The purpose of the experiments is to correlate the wave behaviour with *in situ* measurements of dynamic surface properties rather than the specific chemicals that create these surface properties.

The presence of soluble surfactants in the bulk creates, through adsorption of surfactant molecules, an ambient surface excess concentration of surfactant (Γ_a ; ref. 16). To characterize the dynamical properties of the water surface, the surface pressure isotherm, the equilibrium (Gibbs) surface elasticity and the surface viscosity (the sum of the surface shear and dilational viscosities) were measured *in situ* for each data set (1 to 4b; Fig. 1). The ambient surface tension (σ_a) is plotted against the log of C in Fig. 1a. The data fall on a single straight line with the ambient surface tension decreasing as the bulk concentration increases, as is typical of surfactants¹⁶; however, the measured values are not the same as those found in tests with pure SDS and distilled water¹⁷. The ambient surface tension at the highest SDS concentration was the same for data sets 4a and 4b.

Figure 1b shows plots of the surface pressure isotherms for each data set. The isotherm for data set CLEAN shows a constant surface pressure ($\pi = \sigma_c - \sigma$, where σ_c is the surface tension of clean water and σ is the local surface tension at the measuring device) of zero down to a compression of 75% ($A/A_0 = 0.25$, where A/A_0 is the

normalized surface area in a Langmuir trough that surrounds the surface tension measuring device; see Methods). (It was found that breakers generated in this case behaved in the same manner (see below) as breakers generated in cases with lower values of Γ_a .) In the presence of SDS, the surface pressure increases monotonically from a non-zero initial value as the surface area is compressed. The two surface pressure curves for the highest concentration (data sets 4a and 4b) are the same at low surface compression, but the curve for data set 4b has a much higher slope for $A/A_0 < 0.3$ (−1.2 in the semi-log plot). The slope of these curves is called the static elasticity.

The equilibrium elasticity E_0 and surface viscosity μ_s are plotted against the bulk concentration of SDS in Fig. 1c, d, respectively. Both quantities decrease as the bulk concentration of SDS increases. At the highest bulk concentration, the equilibrium elasticity is about the same on the two measurement days, whereas the surface viscosity shows a marked increase on the last day of the experiment (data set 4b). The measured values of ambient surface pressure and elasticity are within the range of field measurements in some coastal regions^{20,21}.

The histories of the crest shape of the breaking waves were recorded in each data set. As has been reported previously^{9,10,13}, in clean water a bulge forms on the forward face of the wave crest as it steepens and capillary waves form upstream of the leading edge (called the toe) of the bulge (Fig. 2a). (In crest-fixed coordinates, the upstream direction is to the left of the crest.) When the toe becomes sharp (small radius of curvature), the flow separates, the toe begins to move down the wave face (Fig. 2b) and a turbulent flow ensues¹⁵. As the toe moves down the wave face, ripples are generated between the toe and the crest (not shown in Fig. 2), and these ripples propagate backwards over the crest in crest-fixed coordinates. There is no plunging jet formation and impact, or any other overturning of the free surface.

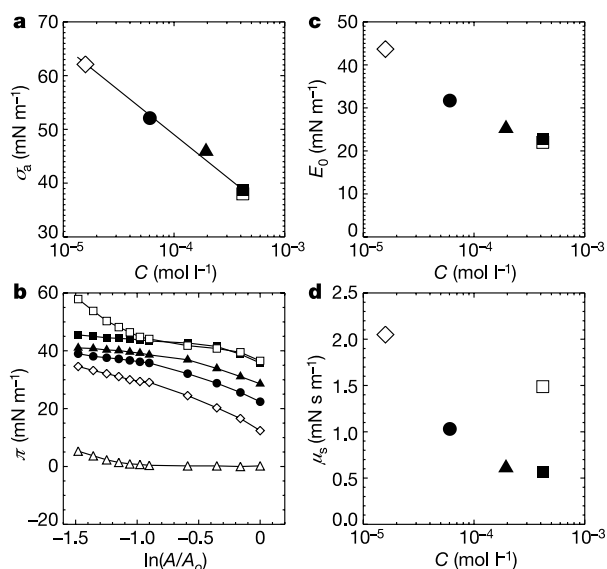


Figure 1 Measurements of dynamic surface properties. Six sets of measurements are shown with various bulk concentrations (C , in mol l⁻¹) of SDS. Open triangles, $C = 0$ (data set CLEAN); open diamonds, $C = 1.6 \times 10^{-5}$ (data set 1); filled circles, $C = 6.0 \times 10^{-5}$ (data set 2); filled triangles, $C = 1.94 \times 10^{-4}$ (data set 3); filled squares, $C = 4.17 \times 10^{-4}$ (data set 4a); open squares, $C = 4.17 \times 10^{-4}$ (data set 4b). Data set 4b was taken 3 d after data set 4a without adding or removing SDS from the tank. **a**, Ambient surface tension, σ_a (mN m⁻¹), versus C . The line is a least squares fit of the experimental data. **b**, Surface pressure, π (mN m⁻¹), versus the log of the normalized water surface area (A/A_0) in a Langmuir trough that surrounds the surface tension measuring device. **c**, Gibbs elasticity, E_0 (mN m⁻¹), versus C . **d**, Surface viscosity, μ_s (mN s m⁻¹), versus C .

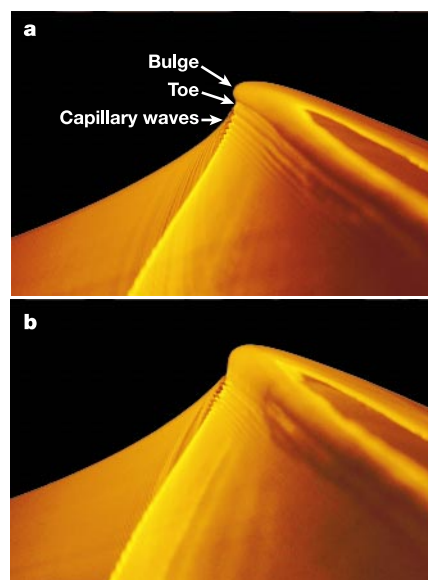


Figure 2 Two images of a breaking wave in 'clean' water. The wave-packet frequency is 1.15 Hz. The wave propagates from right to left. The breaker is highly repeatable and each image is from a different experimental run. The width of each image is 12.5 cm in the plane of the light sheet. In each image, the edge between the upper dark region and the lower wavy light region is the wave crest profile at the centreline of the tank where the laser light sheet intersects the wave surface. The variation of the light intensity below this edge is due to the non-uniform intensity of the underwater portion of the light sheet, which is produced by refraction as it enters the curved water surface and by viewing the glowing dye in the light sheet through the curved water surface between the camera and the light sheet. A detailed description of these effects is given in ref. 13. **a**, The ripple pattern forms at the crest. **b**, The toe begins to move down the wave face.

This process changes in the presence of SDS in several ways. As the concentration of SDS is increased, the capillary waves decrease in amplitude and number, the size of the bulge shrinks, and the ripples generated between the toe and the crest during the spilling process diminish. In the first experiment at the highest SDS concentration (data set 4a), the capillary waves disappear and a bulge forms with a very steep front near the toe. At this point, the toe moves down the wave face in a manner much like that found in the clean water case described above, and the wave breaks without overturning of the free surface.

In the second experiment at the highest SDS concentration (data set 4b), the wave behaviour changed markedly. Four images of one of the breakers are shown in Fig. 3. The wave-maker motion is identical to that for the breaker shown in Fig. 2. The bulge still forms initially, but without a sharp toe and without capillary waves ahead

of the toe (Fig. 3a). The bulge is fairly uniform along the crest (that is, in the direction perpendicular to the plane of the image). After a short time period, a two-dimensional jet with a thickness of about 5 mm emerges from the bulge near the toe (Fig. 3b). The roughness of the jet may be the result of a cross-stream instability found in the jets of plunging breakers in clean water²². The jet ejects forward (Fig. 3c) and re-enters the front face of the wave. As the jet hits the front face of the wave, it entraps a tube of air and a splash projects forward from the impact point (Fig. 3d). The surface motion is more violent than that for a wave generated by the same wave-maker motion in clean water. Images from high-speed movies of the breakers were analysed to obtain the crest profile at each instant. Figure 4a shows the crest profile history of a single breaking wave with an average wave-packet frequency of 1.15 Hz. The jet motion and splashing can be seen more easily from this history record.

Breakers were measured for all three wave-packet frequencies (three waves at 1.42 Hz, three waves at 1.26 Hz, and six waves at 1.15 Hz) in data set 4b. At each frequency, the wave breaks in a different position in the tank. The small jet appears in all of these breaking events and always with an initial thickness of about 5 mm. Figure 4c and d, respectively, show the relative horizontal (x) and vertical (z) distances (that is, relative to the highest point of the wave crest; see Fig. 4b for the coordinate system) of the jet tip (defined as the point where the slope of the water surface relative to the horizontal changes sign) as a function of time (t). It should be mentioned that the jet tip as defined above is not necessarily a material point.

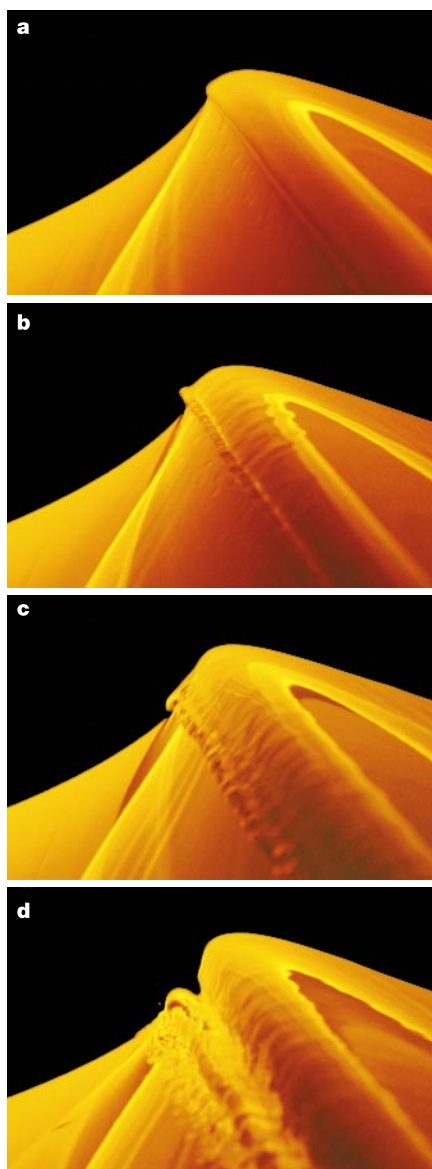


Figure 3 Images of a surfactant-dominated breaking wave generated in a manner identical to the wave in Fig. 2. The breaker is highly repeatable and each image is from a different experimental run. The width of the images is 12.5 cm in the plane of the light sheet. The dynamic properties of the water free surface correspond to data set 4b. Note that the jet does not issue from the highest elevation of the wave as it would in a typical plunging breaker in clean water. **a**, The bulge forms with a weak toe and without capillary waves ahead of the toe. **b**, The jet forms. **c**, The jet follows a parabolic trajectory. **d**, The jet hits the front face of the wave, creating a secondary splash ahead of it.

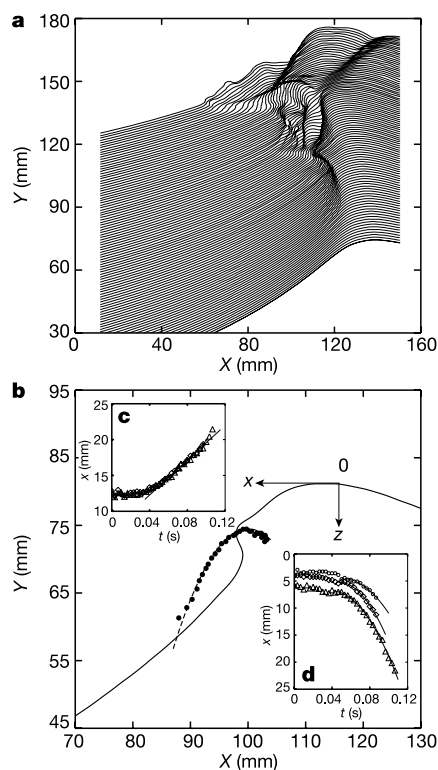


Figure 4 Geometrical properties of the jet. The surfactant conditions correspond to data set 4b. **a**, Crest profile history of a single breaking wave, such as the one shown in Fig. 3. The bottom curve is the first wave profile, and each successive profile is plotted 1.0 mm above the previous one for clarity. The time between each profile is 3.33 ms. **b**, A single wave profile and the trajectory of the tip of the jet. The broken line is a parabola fitted to the experimental data. **c**, **d**, Horizontal and vertical distances, respectively, of the tip of the jet relative to the highest point of the wave crest versus time. The unbroken lines in **d** are parabolas fitted to the data. The wave-maker motion parameters for the three waves shown are $f_0 = 1.42$ Hz (open circles), $f_0 = 1.26$ Hz (open diamonds), $f_0 = 1.15$ Hz (open triangles), where f_0 is the average wave-packet frequency¹³.

The jet position relative to the crest is nearly fixed for the first 0.03 s. The total jet length at impact (~ 0.1 s later) is only ~ 1 cm, whereas the wavelength of the gravity wave is on the order of 1 m. The vertical trajectories (z versus t) are parabolic with an absolute acceleration (that is, relative to the laboratory reference frame) of $6.4 \pm 0.8 \text{ m s}^{-2}$ in all of the waves. The discrepancy between this value and the acceleration of gravity might be due to temporal variations in jet tip shape induced by surface tension. At the instant of impact with the front face of the wave, the absolute vertical velocity of the jet tip is 34.0 ± 1.4 , 25.1 ± 2.5 and $18.9 \pm 3.2 \text{ cm s}^{-1}$ for the wave frequencies of 1.15, 1.26 and 1.42 Hz, respectively. By contrast, the relative horizontal velocity of the jet, which is constant while the vertical trajectory is parabolic, is $10.7 \pm 1.1 \text{ cm s}^{-1}$ at all three frequencies. Numerical simulations of weak breaking events without surface tension²³ have also shown the appearance of a small plunging jet. In the calculations, however, the jet issues from the highest point on the wave.

The measurements indicate that the behaviour of the breaker depends on the various surface dynamic properties in a complex way. Comparison of the wave behaviour and surface properties between data sets 4a and 4b indicates that the appearance of the small plunging jet coincides with high surface viscosity and high static surface elasticity at high surface compression. The difference in either quantity between the two cases would produce a marked change in the distribution of stress in the surfactant film at the crest. The change in elasticity is expected to be important because high surface compression is known to occur at the crests of steep waves^{24,25}. These effects are being explored further by carrying out similar experiments with other soluble surfactants. □

Methods

Water preparation

The procedures for water preparation are as follows. Before the measurements, the tank and piping system were cleaned with highly chlorinated tap water, rinsed and filled with roughly 14,000 l of filtered tap water. The chlorine concentration was increased to 10 p.p.m., and the water was skimmed and filtered for 2 d. Just before the experiments, the chlorine concentration was reduced to nearly zero by adding an appropriate amount of hydrogen peroxide. The fluorescent dye rhodamine 6G (95% purity) was then mixed with the water at a concentration of 0.5 p.p.m. for visualization purposes. Immediately after mixing the dye in the tank, a set of measurements of surface properties and breaker profiles were taken (data set CLEAN). After the measurements, the soluble surfactant SDS (99% purity) was added to the tank water and the measurements were repeated on the following day (data set 1). This procedure (adding SDS at the end of the day and taking measurements on the following day) was repeated three more times over the next 3 d to obtain three additional data sets, 2, 3 and 4a, corresponding to the three increasing SDS concentrations, respectively. At the highest SDS concentration, measurements were also taken 3 d after data set 4a (data set 4b).

It is important to realize that even with the above water treatment procedures, these experiments were not done with solutions of pure water and SDS. The dye and SDS probably contain other surfactants, the water is not distilled, and additional surfactants probably enter the tank by various paths. Also, once the chlorine concentration is reduced to nearly zero, which is done to preserve the dye and SDS, bacteria in the tank begin to grow. After they reach a critical concentration, they increase exponentially in number²⁶. Bacteria are a type of plankton and are known to produce surface-active organic material as by-products of their respiration²⁷. It is speculated that the change in surface properties between data sets 4a and 4b is due to these organic surfactants.

Surface property measurements

Surface property measurements were done with equipment and techniques that are similar to those used in bench-top experiments^{28–30}. The measurement system is centred on an *in situ* Langmuir trough. The trough is essentially a set of Teflon barriers arranged in a rectangular shape (80 cm by 30 cm) that is lowered into the tank to the point where it isolates a small portion of the water surface with area A_0 . One of 30-cm barriers can be moved parallel to the water surface by means of a servo-motor system, thus changing the enclosed area in the trough. The surface pressure isotherms are measured by monitoring the surface tension (measured with a Wilhelmy plate) while compressing the surface area (A) around the plate with the movable barrier over a period of about 2 min. Surface pressure is measured to an accuracy of $\pm 0.2 \text{ mN m}^{-1}$.

The equilibrium (Gibbs) surface elasticity E_0 and surface viscosity μ_s were determined from measurements of the wavelength and decay rate of longitudinal (compressive) surface waves that were generated by oscillating the movable barrier at frequencies ranging from 0.2 to 1.5 Hz. The longitudinal waves were measured with a capillary propagation technique similar to the one described in refs 28–30. The standard errors in the measurements of E_0 and μ_s are $\pm 1.06 \text{ mN m}^{-1}$ and $\pm 0.15 \text{ mN s m}^{-1}$. The measurements indicated that the effect of surfactant diffusion was negligible.

Wave profile measurements

Surface profile measurements were obtained from movies taken at 300 frames s^{-1} with a high-speed black-and-white digital camera (Vision Research, Phantom VI). Each image consists of a 512-by-512 array of pixels with 256 grey levels. The images shown in Figs 2 and 3 were obtained in an identical manner but with the high-speed camera replaced by a high-resolution digital colour camera (Nikon D1x). The illumination for both sets of images consists of a light sheet from an Nd:YAG laser. The light sheet is 1 mm thick at the water surface and oriented vertically, along the centre plane of the tank. The cameras view the intersection of the light sheet and the water surface from the side looking down with an angle of about 5° from the horizontal. A long-pass colour filter was placed in front of the camera lens to prevent the laser light from entering the camera; the light source for the images was the glowing fluorescent dye in the light sheet. The camera and the illumination system were mounted on an instrument carriage that is driven by a servo-motor and controlled by the same computer that controls the wave-maker. The carriage was set to travel along the tank with the crest of the breaking wave so that the movies were taken in crest-fixed coordinates. More detailed descriptions of these techniques are given in ref. 13.

Received 5 June; accepted 9 December 2002; doi:10.1038/nature01357.

1. Cokelet, E. D. Breaking waves. *Nature* **267**, 769–774 (1977).
2. Thorpe, S. A. Small-scale processes in the upper ocean boundary layer. *Nature* **318**, 519–522 (1985).
3. Melville, W. K. The role of surface-wave breaking in air–sea interaction. *Annu. Rev. Fluid Mech.* **28**, 279–321 (1996).
4. Jessup, A. T., Zappa, C. J., Loewen, M. R. & Hesany, V. Infrared remote sensing of breaking waves. *Nature* **385**, 52–55 (1997).
5. Okuda, K. Internal flow structure of short wind waves. Part I. On the internal vorticity structure. *J. Oceanogr. Soc. Jpn* **38**, 313–322 (1982).
6. Ebuchi, N., Kawamura, H. & Toba, Y. Fine structure of laboratory wind–wave surfaces studied using an optical method. *Boundary-Layer Meteorol.* **39**, 133–151 (1987).
7. Longuet-Higgins, M. S. Capillary rollers and bores. *J. Fluid Mech.* **240**, 659–679 (1992).
8. Longuet-Higgins, M. S. Shear instability in spilling breakers. *Proc. R. Soc. Lond. A* **446**, 399–409 (1994).
9. Duncan, J. H., Philomin, H., Behres, H. & Kimmel, J. The formation of a spilling breaker. *Phys. Fluids* **6**, 2558–2560 (1994).
10. Duncan, J. H., Philomin, V., Qiao, H. & Kimmel, J. The formation of a spilling breaker. *Phys. Fluids* **6**, S2 (1994).
11. Longuet-Higgins, M. S. Capillary jumps on deep water. *J. Phys. Oceanogr.* **29**, 1957–1965 (1996).
12. Tulin, M. P. in *Waves and Nonlinear Processes in Hydrodynamics* (eds Grue, J., Gjevik, B. & Weber, J. E.) 177–190 (Kluwer, Dordrecht, 1996).
13. Duncan, J. H., Qiao, H., Philomin, V. & Wenz, A. Gentle spilling breakers: crest profile evolution. *J. Fluid Mech.* **379**, 191–222 (1999).
14. Cenicerros, H. D. & Hou, T. Y. Dynamic generation of capillary waves. *Phys. Fluids* **11**, 1032–1050 (1999).
15. Qiao, H. & Duncan, J. H. Gentle spilling breakers: crest flow-field evolution. *J. Fluid Mech.* **439**, 57–85 (2001).
16. Edwards, D. A., Brenner, H. & Wasan, D. T. *Interfacial Transport Process and Rheology* 1st edn (Butterworth-Heinemann, Boston, Massachusetts, 1991).
17. Poskanzer, A. M. & Goodrich, F. C. Surface viscosity of sodium dodecyl sulfate solutions with and without added dodecanol. *J. Phys. Chem.* **79**, 2122–2126 (1975).
18. Longuet-Higgins, M. S. *Proc. 10th Symp. Naval Hydrodynamics* 597–605 (US Government Printing Office, Arlington, Virginia, 1976).
19. Melville, W. K. & Rapp, R. J. Momentum flux in breaking waves. *Nature* **317**, 514–516 (1985).
20. Wei, Y. & Wu, J. *In situ* measurements of surface tension, wave damping, and wind properties modified by natural films. *J. Geophys. Res.* **97**, 5307–5313 (1992).
21. Barger, W. R., Daniel, W. H. & Garrett, W. D. Surface chemical properties of banded sea slicks. *Deep-Sea Res.* **21**, 83–89 (1974).
22. Longuet-Higgins, M. S. On the disintegration of the jet in a plunging breaker. *J. Phys. Oceanogr.* **25**, 2458–2462 (1995).
23. Longuet-Higgins, M. S. & Dommermuth, D. G. Crest instabilities of gravity waves. Part 3. Nonlinear development and breaking. *J. Fluid Mech.* **336**, 33–50 (1997).
24. Longuet-Higgins, M. S. & Cokelet, E. D. The deformation of steep waves on water. I. A numerical method of computation. *Proc. R. Soc. Lond. A* **350**, 1–26 (1976).
25. Cenicerros, H. D. The effects of surfactants on the formation and evolution of capillary waves. *Phys. Fluids* **15**, 245–256 (2003).
26. Sokatch, J. R. *Bacterial Physiology and Metabolism* (Academic, London, 1969).
27. Frew, N. M. in *The Sea Surface and Global Change* (eds Liss, P. S. & Duce, R. A.) 121–172 (Cambridge Univ. Press, Cambridge, 1997).
28. Miyano, K., Abraham, B. M., Ting, L. & Wasan, D. T. Longitudinal surface waves for the study of dynamic properties of surfactant systems. I. Instrumentation. *J. Colloid Interface Sci.* **92**, 297–302 (1983).
29. Ting, L., Wasan, D. T., Miyano, K. & Xu, S.-Q. Longitudinal surface waves for the study of dynamic properties of surfactant systems. II. Air–solution interface. *J. Colloid Interface Sci.* **102**, 248–259 (1984).
30. Mass, J. T. & Milgram, J. H. Dynamic behavior of natural sea surfactant films. *J. Geophys. Res.* **103**, 15695–15715 (1998).

Acknowledgements We thank H. Qiao for preliminary experiments on breaking waves in the presence of surfactants, and G. M. Korenowski for conversations on surfactant chemistry. This work is supported by the Ocean Sciences Division of the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.H.D. (e-mail: duncan@eng.umd.edu).

Damage to the shallow Landers fault from the nearby Hector Mine earthquake

John E. Vidale* & Yong-Gang Li†

* Earth & Space Sciences Department and Institute of Geophysics & Planetary Physics, University of California in Los Angeles, Los Angeles, California 90095-1567, USA

† Department of Earth Sciences, University of Southern California, Los Angeles, California 90089-0740, USA

Crustal faults have long been identified as sites where localized sliding motion occurs during earthquakes¹, which allows for the relative motion between adjacent crustal blocks. Although there is a growing awareness that we must understand the evolution of fault systems on many timescales to relate present-day crustal stresses and fault motions to geological structures formed in the past, fault-zone damage and healing have been documented quantitatively in only a few cases^{2,3}. We have been monitoring the healing of damage on the shallow Johnson Valley fault after its rupture in the 1992 magnitude-7.3 Landers earthquake, and here we report that this healing was interrupted in 1999 by the magnitude-7.1 Hector Mine earthquake rupture, which occurred 20–30 km away. The Hector Mine earthquake both strongly shook and permanently strained the Johnson Valley fault, adding damage discernible as a temporary reversal of the healing process. The fault has since resumed the trend of strength recovery that it showed after the Landers earthquake. These observations lead us to speculate that fault damage caused by strong seismic waves may help to explain earthquake clustering and seismicity triggering by shaking, and may be involved in friction reduction during faulting.

Our field experiments (Fig. 1) were conducted in 1994, 1996, 1997, 1998, 2000 and 2001. The 1996 and 1997 experiments were less comprehensive owing to logistical difficulties. Each experiment was very similar: seismometers were re-buried in the same holes and the explosions were in 35-m drill holes, which were sited within 10 m of the previous years' holes. The seismic arrays line 1 and line 3 had 36 and 21 seismometers, respectively. The first repetition was used to map Landers fault-zone fine structure as a function of depth⁴. With the added iterations, we have watched the fault evolve with time.

We infer that the rupture reduced the seismic wave speeds, which have been subsequently recovering, although data from before the Landers earthquake are lacking. Figure 2a–c shows the increase of compressional (P)-wave and shear (S)-wave velocity with time after the 1992 Landers rupture. The 1994 and 1996 observations of healing at Landers showed recovery of seismic velocity by about 2%. The survey in 1998 showed a reduction of the healing rate by a factor of two between 1994–96 and 1996–98. The ratio of the rates of P-wave and S-wave speed recovery is consistent with healing caused by closure of cracks that are partially fluid-filled⁵. A similar experiment at Hector Mine has confirmed that healing is not unique to Landers and shows that there is variability in healing rates among the fault segments that we have measured³. Our results are similar to an observed drop in wave speed at the time of the Loma Prieta earthquake, which recovered over subsequent years^{6,7} and has also shown damage from a smaller event³. By contrast, another study does not show healing 5–7 yr after the Kobe earthquake, but does suggest damage and rapid healing on the Kobe fault zone resulting from the Tottori earthquake⁸.

The velocity increase is consistent and diminishing in rate over time across the 1994, 1996, 1997 and 1998 surveys. In the 2000

survey, however, the velocities in the fault zone decreased for both types of wave and all six ray paths, although the velocities for the outer stations did not significantly change. The following year (2001), the survey showed a resumption of healing for both types of wave on all lines. Figure 3 shows the depth dependence of the inferred healing.

The magnitude-7.1 1999 Hector Mine earthquake occurred between the 1998 and 2000 surveys, and provides the most likely cause for the temporary reversal in the polarity of the velocity change. An alternative is that the water table was tens of metres higher in 2000 than in other years, but this is unlikely because water table changes of that magnitude would be surprising. In addition, water table changes would affect the P-wave velocity much more than the S-wave velocity, contrary to the change observed, and 1997 and 2001 are the years that had the greatest rainfall (see Western Regional Climate Center website: <http://www.wrcc.dri.edu/>).

The Hector Mine earthquake could have caused the observed change in seismic wave speed in two ways. These involve dynamic and static stresses, which have already been cited as possible mechanisms for triggering distant aftershocks and inducing deformation on faults. Dynamic stresses during the Hector Mine earthquake on the Johnson Valley fault were a few megapascals, and static stress changes caused by the earthquake are perhaps half a megapascal (ref. 9).

Most probably, the dynamic stresses during the strong shaking cracked connections in the rock—as we infer the Landers mainshock did in 1992—but to a lesser degree. This damage is healing in subsequent years, just as we observe the main shock damage to be healing. Such shaking-induced damage has been observed in laboratory studies¹⁰. The concept of our model is shown in Fig. 4.

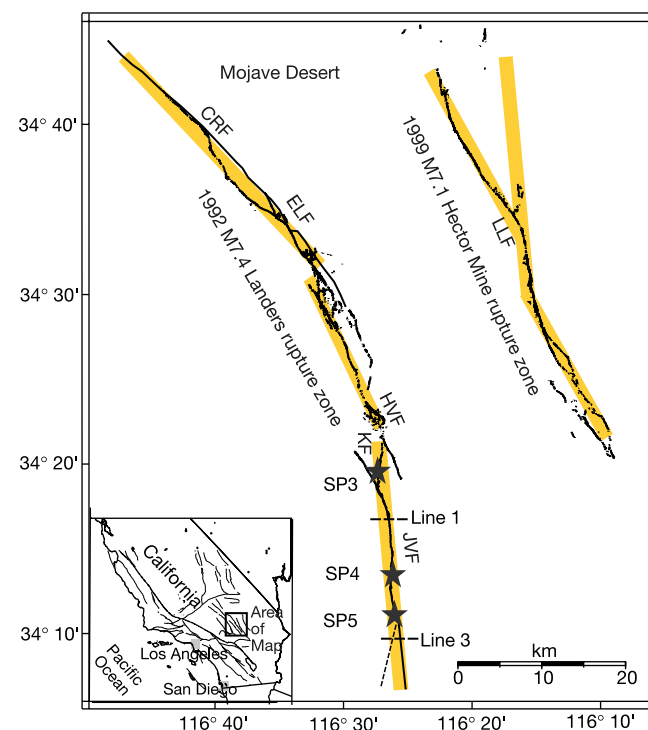


Figure 1 Map of Landers and Hector Mine ruptures. Black lines indicate mapped surface breakage^{25,26}, yellow lines show rupture inferred from seismic and geodetic studies^{27,28}, stars show the locations of shot points SP3, SP4 and SP5, and broken lines show the extent of the seismometer profiles line 1 and line 3, which have 36 and 21 seismometers, respectively. Fault segments are labelled as follows: CRF, Camp Rock fault; ELF, Emerson Lake fault; HVF, Homestead Valley fault; JVF, Johnson Valley fault; KF, Kickapoo fault; LLF, Lavic Lake fault.

We also examined seismicity in the study area. At the time of the Hector Mine earthquake the seismicity increased, which is consistent with weakening from shaking-induced damage. The spatial pattern also changed, however, precluding a simple interpretation.

Alternatively, the static stress change caused by the Hector Mine slip might change the crack density and orientation around Landers, much as it probably strained the fault zone according to an InSAR study⁹. This explanation does not work as well as the shaking from dynamic waves for several reasons, although it remains viable. It is not clear why the velocity would be reduced for both P and S waves: stresses could lead to either closure or opening of cracks, depending on their orientation. Also, the strain observed around the faults was a step function in time⁹, whereas the velocity changes that we report were more protracted. In addition, we have not observed obvious changes in the S-wave splitting directions or amplitudes in surveys after the Landers and Hector earthquakes as healing progressed, as might be expected from systematic changes in the crack population¹¹.

However, velocity changes from stress changes have been suggested previously^{12,13}, and we cannot conclusively rule out that static stress change is the cause of the observed damage. The observation of strain concentrated on fault zones⁹ is not inconsistent with our interpretation of dynamic shaking as the cause of the damage, because concentrated deformation may help to cause damage, and low-strength fault zones are doubly susceptible to damage because of their fractured nature and the wave amplification caused by their low impedance. It is also possible that several mechanisms contributed to the velocity reduction.

If the observed change in seismic velocity is modelled by fluid-filled cracks, then the degree of fluid saturation can be estimated from the ratio between the P-wave and S-wave travel-time changes. For example, S-wave velocity does not depend as strongly as P-wave velocity on the amount of water in cracks. Figure 2d shows this ratio, which suggests that the change in velocity can be explained by closure of partially saturated cracks. We have inferred a trend of

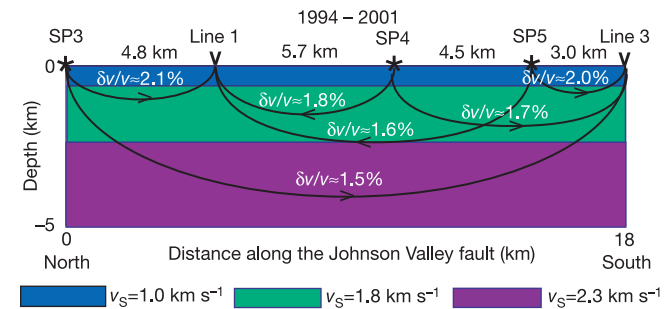


Figure 3 Model of depth dependence of ground healing. Shown are the approximate depth of seismic ray penetration and the inferred accumulated healing at each depth.

progressively wetter cracks with time from the evolution of the ratios⁵, and this trend is continued in the new data for 2000 and 2001. If this model applies, the degree of fluid saturation has risen from 60 to 80%, and this rise may be caused by ground water percolating into the cracks that opened in the Landers earthquake in combination with diminishing crack pore volume. We assumed isotropically oriented cracks in our analysis, given the limited range of ray paths in our experiment, although the orientation of the closing cracks may be anisotropic.

Earthquake activity has been seen to be more sensitive to regional large earthquakes than was previously thought. Aftershocks can be triggered up to 500 km from the rupture plane¹⁴, even in non-volcanic areas¹⁵. Changes in seismicity out to distances of ten rupture lengths have been proposed to indicate impending earthquakes^{16,17}. Earthquakes cause sympathetic slip on nearby faults^{18,19}. Deformation from stresses of nearby earthquakes may show that faults are much more compliant than the surrounding harder rock on the scale of 1–2 km (ref. 9). Reduced strength has also been

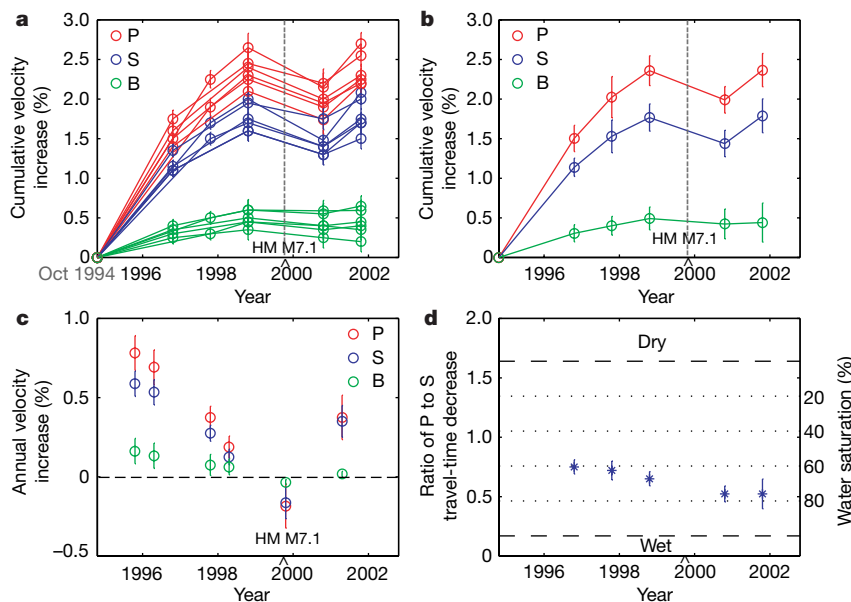


Figure 2 Velocity measurement of the Landers fault from 1994 to 2001. **a**, Change in P-wave and S-wave velocity for each of the six ray paths (three shots recorded by two arrays) along the fault measured at the five stations nearest the surface fault break. Four stations outside the fault zone, about 1 km on either side of the surface fault trace (B, green symbols), show less healing than do stations nearer the fault trace. **b**, Average and standard deviation from combining the six paths. **c**, The estimated annual increase in velocity, which is the derivative of the curves in **b**. **d**, Ratio of P-wave travel-time changes

to S-wave travel-time changes, which is sensitive to the degree of fluid saturation in cracks. Horizontal lines indicate the ratios predicted for a range of water saturation percentages. Vertical grey lines in **a** and **b** indicate the time of the Hector Mine (HM) earthquake (16 October 1999). Experiments took place in October of each year. The stations outside the fault zone are W16, W15, E15 and E14, and the stations near the fault break are W2, W1, ST0, E1 and E2 (ref. 5).

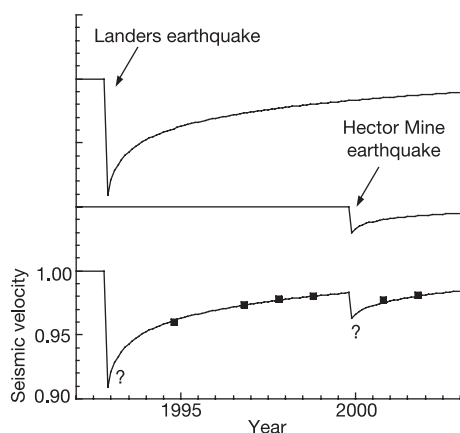


Figure 4 Model of velocity as a function of time owing to damage from the Landers rupture, the Hector Mine shaking, and the combination of the two compared with observations. Shown is healing as a logarithm of time¹⁰, although details just after each event and extrapolating into the future are not well constrained. The velocity before the Landers earthquake was not measured.

inferred from fault-zone-guided waves on the scale of a few hundred metres⁴.

Previously we showed that fault zones heal⁵, and here we have shown that the combination of shaking and static stress reduces the modulus of a recently broken fault zone. This probably indicates that the strength of the fault was reduced, which could trigger earthquakes. Our observation thus may provide a direct measurement of the missing connection between shaking and the facilitation of distant aftershocks^{20,21}. Regional clustering of earthquakes is another plausible result of one big earthquake weakening the regional set of faults around it, either as an anomalous activation of a region²² or as the progression of ruptures along a fault, with the shaking modulated by directivity²³.

Another implication of our result is that existing friction laws may need improvement. Currently, friction is modelled simply (or not so simply) as a function of a state variable, which is the history of sliding, and current sliding rate of a point on a fault plane²⁴. If shaking can significantly reduce strength, it may also help to explain the puzzle of aftershock occurrence very near the mainshock fault plane, which strikes where the Earth has been strongly shaken but where the regional stress is reduced. Shaking-induced weakening also may be involved in the progression of rupture during earthquakes, because strong shaking is likely to precede the arrival of the rupture front. □

Received 22 August; accepted 29 November 2002; doi:10.1038/nature01354.

- Reid, H. F. in *The California Earthquake of April 18, 1906: Report of the State Earthquake Investigation Commission* (Carnegie Institution of Washington, Washington DC, 1910).
- Li, Y.-G., Vidale, J. E., Day, S. M., Oglesby, D. D. & Cochran, E. Post-seismic fault healing on the rupture zone of the 1999 M7.1 Hector Mine, California earthquake. *Bull. Seismol. Soc. Am.* (in the press).
- Rubinstein, J. L., Beroza, G. C., Bokelmann, G. & Schaff, D. Near surface damage caused by the strong ground motion of the M6.9 Loma Prieta and M5.4 Chittenden earthquakes. *Eos* **83**, NG21B (2002).
- Li, Y.-G., Vidale, J. E., Aki, K., Xu, F. & Burdette, T. Depth-dependence structure of the Landers fault zone from trapped waves generated by aftershocks. *J. Geophys. Res.* **105**, 6237–6254 (2000).
- Li, Y.-G. & Vidale, J. E. Healing of the shallow fault zone from 1994–1998 after the M7.5 Landers, California, earthquake. *Geophys. Res. Lett.* **28**, 2999–3002 (2001).
- Schaff, D. P. & Beroza, G. C. Postseismic response of repeating aftershocks. *Geophys. Res. Lett.* **25**, 4549–4552 (1998).
- Baisch, S. & Bokelmann, G. H. R. Seismic waveform attributes before and after the Loma Prieta earthquake: scattering change near the earthquake and temporal recovery. *J. Geophys. Res.* **106**, 16323–16337 (2001).
- Ikuta, R., Yamaoka, K., Miyakawa, K., Kunitomo, T. & Kumazawa, M. Continuous monitoring of propagation velocity of seismic wave using ACROSS. *Geophys. Res. Lett.* **29**, 5-1–5-4 (2002).
- Fialko, Y. *et al.* Deformation on nearby faults induced by the 1999 Hector Mine earthquake. *Science* **297**, 1858–1862 (2002).
- Richardson, E. & Marone, C. Effects of normal stress vibrations on frictional healing. *J. Geophys. Res.* **104**, 28859–28878 (1999).
- Cochran, E. & Vidale, J. E. Seismic anisotropy of the Hector Mine fault zone. *Eos* **82**, S41A (2001).

- Bokelmann, G. H. R. & Harjes, H.-P. Evidence for temporal variation of seismic velocity within the upper continental crust. *J. Geophys. Res.* **105**, 23879–23894 (2000).
- Ratdomopurbo, A. & Poupinet, G. Monitoring a temporal change of seismic velocity in a volcano: Application to the 1992 eruption of Mt. Merapi (Indonesia). *Geophys. Res. Lett.* **22**, 775–778 (1995).
- Hill, D. P. *et al.* Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* **260**, 1617–1623 (1993).
- Brodsky, E., Karakostas, V. & Kanamori, H. A new observation of dynamically triggered regional seismicity: earthquakes in Greece following the August, 1999 Izmit, Turkey earthquake. *Geophys. Res. Lett.* **27**, 2741–2744 (2000).
- Knopoff, L., Levshina, T., Keilis-Borok, V. I. & Mattoni, C. Increased long-range intermediate-magnitude earthquake activity prior to strong earthquakes in California. *J. Geophys. Res.* **101**, 5779–5796 (1996).
- Bowman, D. D. & King, G. G. C. P. Accelerating seismicity and stress accumulation before large earthquakes. *Geophys. Res. Lett.* **28**, 4039–4042 (2001).
- Price, E. & Sandwell, D. T. Small-scale deformations associated with the 1992 Landers, California, earthquake mapped by synthetic aperture radar interferometry. *J. Geophys. Res.* **103**, 27001–27016 (1998).
- Fialko, Y., Simons, M. & Agnew, D. The complete (3-D) surface displacement field in the epicentral region of the 1999 Mw7.1 Hector Mine earthquake, southern California, from space geodetic observations. *Geophys. Res. Lett.* **28**, 3063–3066 (2001).
- Kilb, D., Gomberg, J. & Bodin, P. Triggering of earthquake aftershocks by dynamic stresses. *Nature* **408**, 570–574 (2000).
- Gomberg, J., Reasenberg, P. A., Bodin, P. & Harris, R. A. Earthquake triggering by seismic waves following the Landers and Hector Mine earthquakes. *Nature* **411**, 462–466 (2001).
- Rockwell, T. K. *et al.* Paleoseismology of the Johnson Valley, Kickapoo, and Homestead Valley faults: clustering of earthquakes in the eastern California shear zone. *Bull. Seismol. Soc. Am.* **90**, 1200–1236 (2000).
- Stein, R. S. Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophys. J. Int.* **128**, 594–604 (1997).
- Marone, C. Laboratory-derived friction laws and their application to seismic faulting. *Annu. Rev. Earth Planet. Sci.* **26**, 643–696 (1998).
- Sieh, K. *et al.* Near-field investigation of the Landers earthquake sequence, April to July 1992. *Science* **260**, 171–176 (1993).
- Scientists of the USGS, SCEC, and CDMG. Preliminary report on the 16 October 1999 M7.1 Hector Mine, California, earthquake. *Seismol. Res. Lett.* **71**, 11–23 (2000).
- Hauksson, E., Jones, L. M. & Hutton, K. The 1999 Mw7.1 Hector Mine, California earthquake sequence: complex conjugate strike-slip faulting. *Bull. Seismol. Soc. Am.* **92**, 1154–1170 (2002).
- Ji, C., Wald, D. J. & Helmberger, D. V. Source description of the 1999 Hector Mine, California earthquake. Part II: complexity of slip history. *Bull. Seismol. Soc. Am.* **92**, 1208–1226 (2002).

Acknowledgements We thank E. Hauksson for supplying relocations of regional seismicity; H. Houston, P. Davis, E. Brodsky, C. Marone, R. Stein and L. Knopoff for comments; and T. Burdette, F. Xu and E. Cochran for field help.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.E.V. (e-mail: vidale@ucla.edu).

Head and backbone of the Early Cambrian vertebrate *Haikouichthys*

D.-G. Shu^{††}, S. Conway Morris[‡], J. Han^{*}, Z.-F. Zhang^{*}, K. Yasui[§], P. Janvier^{||}, L. Chen^{*}, X.-L. Zhang^{*}, J.-N. Liu^{*}, Y. Li^{*} & H.-Q. Liu^{*}

^{*} Early Life Institute and Department of Geology, Northwest University, Xi'an, 710069, China

[†] School of Earth Sciences and Resources, China University of Geosciences, Beijing, 100083, China

[‡] Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

[§] Institute of Molecular Embryology and Genetics, Division of Development and Biohistory, Kumamoto University, 2-2-1 Honjo, Kumamoto 860-0811, Japan
^{||} UMR 8569 du CNRS, Laboratoire de Paléontologie, Muséum National d'Histoire Naturelle, 8 rue Buffon, Paris 75005, France, and Department of Palaeontology, Natural History Museum, London SW7 5BD, UK

Agnathan fish hold a key position in vertebrate evolution, especially regarding the origin of the head and neural-crest-derived tissue¹. In contrast to amphioxus², lampreys and other vertebrates possess a complex brain and placodes that contribute to well-developed eyes, as well as auditory and olfactory systems³.

These sensory systems were arguably a trigger to subsequent vertebrate diversifications. However, although they are known from skeletal impressions in younger Palaeozoic agnathans⁴, information about the earliest records of these systems has been largely wanting. Here we report numerous specimens of the Lower Cambrian vertebrate *Haikouichthys ercaicunensis*, until now only known from the holotype⁵. *Haikouichthys* shows significant differences from other fossil agnathans: key features include a small lobate extension to the head, with eyes and possible nasal sacs, as well as what may be otic capsules. A notochord with separate vertebral elements is also identifiable. Phylogenetic analysis indicates that this fish lies within the stem-group craniates. Although *Haikouichthys* somewhat resembles the ammocoete larva of modern lampreys, this is because of shared general craniate characters; adult lampreys and hagfishes (the cyclostomes if monophyletic^{6,7}) are probably derived in many respects.

The putative Early Cambrian agnathan *Haikouichthys*, from the Chengjiang Lagerstätte⁸ near Kunming, Yunnan, was only known from a single, incomplete specimen. Although its vertebrate affinities have been widely accepted^{9,10}, many of the conclusions about its anatomy and thereby its phylogenetic position are regarded as 'highly provisional'¹¹. The discovery of more than 500 specimens, from a locality near Haikou, reveals a series of new and unexpected features that imply a major reconsideration of several features of early agnathan evolution.

Most notable in this respect is the identification in at least 300 specimens of a small (usually less than 1 mm in length) lobate extension at the anterior end of the animal (Fig. 1a–d). It is separated from the rest of the head by a slight constriction. Its most conspicuous feature is a pair of dark oval stains, interpreted as eyes. Although an approximately circular area in some specimens may indicate the lens, this feature is too inconsistent to be reliable. In a few specimens (Fig. 1e, g) the position of an eye is marked as a concave impression, consistent with it having a scleral layer. In life, the eyes may have been relatively flat, although the quality of preservation does not allow inferences on the presence of extra-ocular muscles. The upward direction of the eyes and their position on the anterior lobe suggests, however, that their mobility may have been restricted. Between the eyes is a median structure, often paired and preserved in a similar manner to the eye stains. This may represent part of the olfactory organ, possibly the nasal sacs (Fig. 1a–d). An alternative interpretation, that these structures represent the pineal/parapineal complex¹², is considered less likely. This is because of their relative sizes, position with respect to the eyes, paired nature, and style of preservation suggesting relatively tough tissue. Along the anterior edge of this lobate extension to the head there is a pair of arcuate structures, meeting about a notch that may define the position of a median nostril (Fig. 1a–d). Typically the arcuate structures have relief, and in life were apparently plate-like and possibly composed of cartilage. Other than the plates, no other part of the anterior lobe appears to have been reinforced. More posteriorly, however, the next region of the head is often heavily pervaded by diagenetic mineralization, probably originally sulphides but now oxidized. In the holotype an attempt was made to reconcile this mineralization with the organization of the cranial cartilages seen in the lampreys⁵. The distribution of this mineralization in the new material shows considerable variation and makes this comparison problematical. There is, however, some consistency in the expression of a pair of sub-circular areas, and these may represent the otic capsules (Fig. 1a, b), although no internal structure is evident.

The mouth is not clearly visible, but a ventral recessed area immediately behind the anterior lobe may indicate the position of the oral opening. Another newly observed feature is a series of more or less square-shaped structures extending posteriorly within the main dorsal region of the head and anterior trunk. These are

sometimes connected by a broad, dark strand. This feature is interpreted as a series of separated vertebral elements (arcualia), associated with the notochord (Fig. 1e–l). The shape of the individual vertebral elements is somewhat variable. They may be bifid or arched, and in some cases appear to have encompassed the entire notochord. Up to ten such vertebral elements have been identified, and it is likely that the series extended further backwards, but it is obscured by the overlying impressions of the trunk myomeres. The original composition was possibly of cartilage, and occasional diagenetic mineralization may reflect some degree of calcification. The last noteworthy feature that is not apparent in the holotype is lamellate regions located between the branchial supports on the lower side of the head region. These most probably represent the gills, and on occasion are delimited by boundaries that suggest an original pouch-like structure (Fig. 1e, g, i–l). The entire branchial area generally has a much rougher texture than the remainder of the head, from which it is clearly delimited.

In other respects the new material of *Haikouichthys* largely confirms the earlier observations⁵. The more complete specimens suggest that the posterior part of the trunk tapered evenly (Fig. 1f, h, j, l). Although preservation of the caudal region is poor, there is no evidence for a caudal fin. The presence of a dorsal fin and what appears to be a ventral fin-fold is confirmed, although unequivocal evidence of the latter being paired is not available (Fig. 1f, h, j, l). In the dorsal fin the associated fin-rays are only occasionally visible, perhaps owing to the original thickness of the dorsal fin, but as in the holotype they appear to be anteriorly tilted. The myosepta are well defined, and at the posteriorly directed flexure of the myomeres some specimens show repeated areas of positive relief. Additional details are not visible, but if these correspond to a set of underlying gonads⁵ (Fig. 1j, l) then this may indicate a primitive metameric arrangement reminiscent of amphioxus. A dark strand running posteriorly on the lower side of the trunk may represent the intestine. There is some evidence that the anus was sub-terminal, thus defining a short post-anal region (Fig. 1f, h, j, l). In the head region the branchial arches are confirmed as a series of posteriorly recurved structures, each apparently composed of a single unit (Fig. 1e, g, i, k). The exact total is equivocal because of variable preservation, but it is most probably seven or eight.

Haikouichthys was evidently a swimmer, although its degree of activity is conjectural. We note, however, co-occurrence of this fish with other pelagic taxa such as *Xidazoon*, and the relative scarcity of benthic forms. The specimens may have been buried alive, possibly as a result of storm-induced burial. Most specimens were collected from a series of graded beds consisting of a lower sand/siltstone unit (about 3 mm thick) and an overlying mudstone, up to 50 mm thick. The typical occurrence is close to the boundary between the two units, and 'shoals' of specimens often show a preferred orientation. Specimens tend to occur in bed-parallel orientation, but where buried obliquely the portion in the mudstone is usually much better preserved than that in the adjacent silty unit.

The new material of *Haikouichthys* shows that our knowledge of the earliest agnathans has been incomplete, with the implication that in several respects the cyclostomes, although almost unchanged since the Carboniferous¹³, are in fact much derived. In particular, the anterior lobe with its eyes (and possible olfactory organ) has no direct counterpart in the other early craniates. There is, however, a possible similarity in the anterior position of the prominent eyes of conodonts¹⁴ and the Ordovician arandaspids¹⁵, although in the latter group this feature has been interpreted as a specialization¹⁶. Anteriorly located eyes are also known in some of the 'naked' fossil agnathans, most notably *Jamoytius*¹⁷ and probably also the anaspid-like fish *Euphanerops*⁴. If this condition of anterior eyes is primitive to vertebrates, it may indicate derivation from the frontal eye of an amphioxus-like ancestor¹⁸, although this would entail various changes including their location on a stable platform and the development of a balancing mechanism¹⁹, which in the form of a

correlated vestibulo-ocular system represents a key step in vertebrate evolution²⁰. It also suggests that the more usual posterior position of the eyes is a result of rostral growth, perhaps in response to the increase in size of the olfactory organ. The more tenuous identification of the olfactory organ in *Haikouichthys* makes its phylogenetic role more speculative, but if correctly identified its

position and size is more comparable to that seen in lampreys, and to some extent hagfish, including fossil representatives¹³, than in most other agnathans and the gnathostomes. The branchial arches of *Haikouichthys* are unlike the intricate branchial basket of lampreys. Their apparently simple structure is somewhat reminiscent of the arrangement seen in the gnathostomes, especially regarding

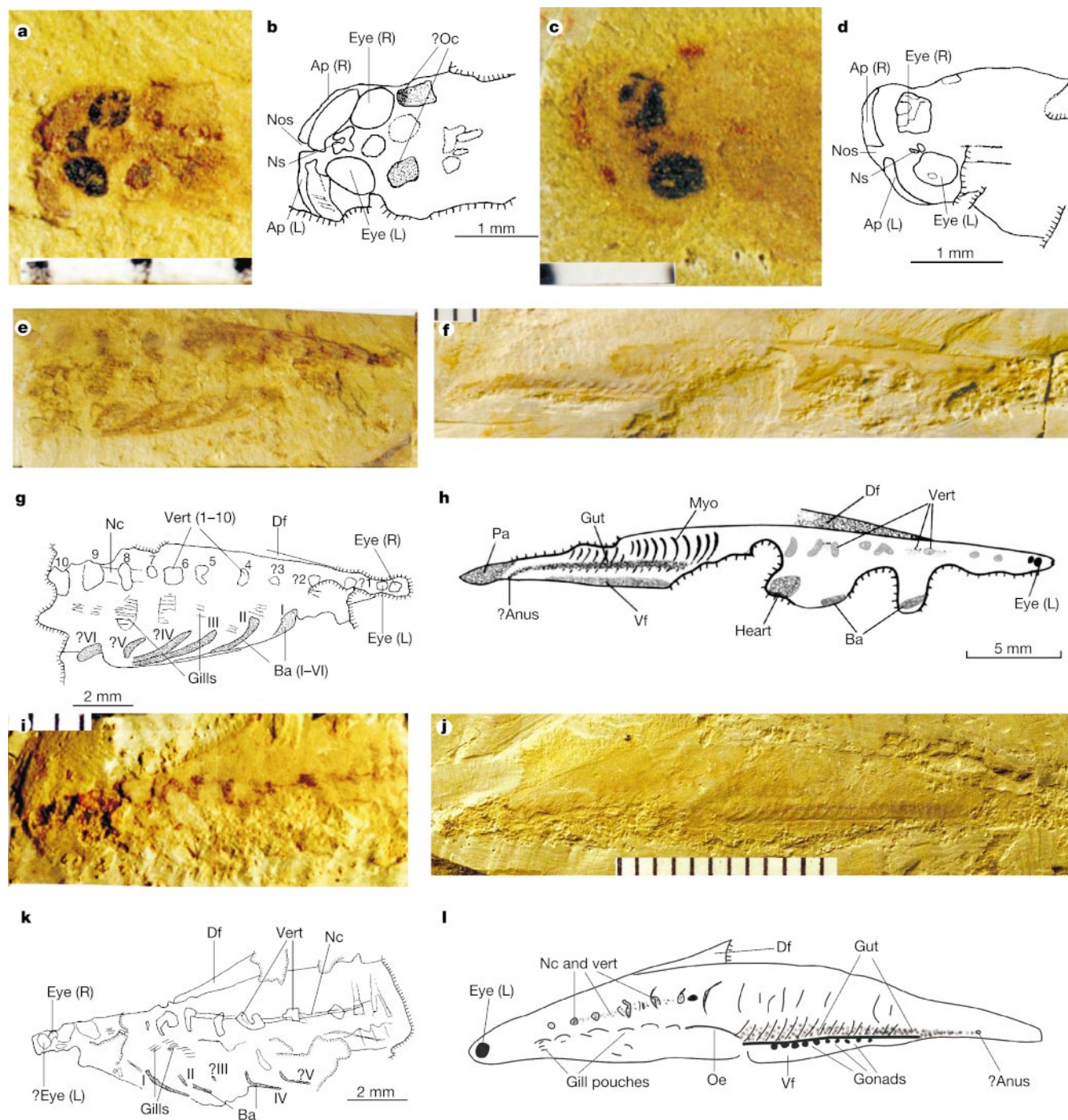


Figure 1 *Haikouichthys ercaicunensis* from Haikou, Kunming, Yunnan. **a–d**, Details of the head, emphasizing its anterior structures; preservation is dorso-ventral. **a, b**, ELI-0001003 (273). **c, d**, ELI-00010013 (323). **e, g, i, k**, Details of the anterior, emphasizing the notochord with associated vertebral elements and branchial arches. **e, g**, ELI-0001015 (12B), anterior to the right. **i, k**, ELI-0001020 (8), anterior end to the

left. **f, h, j, l**, Nearly complete specimens. **f, h**, ELI-0001002 (191), anterior end to the right. **j, l**, ELI-0001001 (172), anterior end to the left. Ap, anterior plates; Ba, branchial arches; Df, dorsal fin; Myo, myosepta; Nc, notochord; Nc and vert, notochord with vertebral elements; Nos, nostril; Ns, nasal sacs; ?Oc, otic capsule; Oe, oesophagus; Pa, post-anal tail; Vert, vertebral elements; Vf, ventral fin-fold. L, left; R, right.

lower units known as the ceratobranchials and hypobranchials. This suggests that the arrangement of articulated branchial elements in gnathostomes retains some primitive characters, even though it is clear that the origin of the jaw entailed developmental rearrangements^{21–23}. The widely spaced vertebral blocks probably acted as supportive structures for the notochord, and as such are reminiscent of the arcualia of adult lampreys. In *Haikouichthys*, however, the vertebral units are larger and more regularly spaced.

The possession of eyes (and probably nasal sacs) is consistent with *Haikouichthys* being a craniate, indicating that vertebrate evolution was well advanced by the Early Cambrian. Although evidently a jawless fish, its precise phylogenetic position is still speculative because this fish shows a puzzling mixture of characters contrary to some previous expectations. Nevertheless, several features of *Haikouichthys*, including what may be metamerically arranged gonads and the anteriorly located eyes, suggest that 'the first fish' may be best regarded as a stem-group craniate (Fig. 2a). Specific connections to other agnathans include the vertebral elements and probable nasal sacs. Despite some similarities to lampreys, in certain respects these living agnathans are probably highly derived.

Until now the other Chengjiang agnathan, *Mylokunmingia*, is only known from a single specimen recovered from a separate locality at a slightly higher stratigraphic level within the Qiongzhusi Formation⁵. Its most obvious differences from *Haikouichthys* are possession of fewer gill pouches (five or six), absence of curved branchial supports, the presence of possible exhalant chambers, and a more anterior extension of the dorsal fin that lacks obvious fin rays. It is clear that the Chengjiang Lagerstätte offers unique insights into early deuterostome diversifications^{5,24–26}, and continued

excavations are expected to extend further our knowledge of their earliest evolution, including that of the vertebrates. □

Methods

Phylogenetic analysis

The analysis, based on the data of ref. 27, was done by using the phylogenetic package HENNIG86 1.5 (ref. 28) and the matrix and tree editor TREE GARDENER (ref. 29) with a data matrix of 17 taxa (4 extant, 13 fossil, cephalochordates as the outgroup) and 115 morphological and physiological characters (see Supplementary Information). All characters are coded binarily as present/absent. Non-applicable characters are coded as 0 and missing data are coded as '?'. Equally, most parsimonious trees were obtained by using the implicit enumeration* (ie*) command.

Received 18 June; accepted 18 October 2002; doi:10.1038/nature01264.

- Northcutt, R. G. The origin of craniates: neural crest, neurogenic placodes, and homeobox genes. *Isr. J. Zool.* **42**, S273–S313 (1996).
- Holland, L. Z. & Holland, N. D. Evolution of neural crest and placodes: amphioxus as a model for the ancestral vertebrate? *J. Anat.* **199**, 85–98 (2001).
- Kleerekoper, H. *The Biology of Lampreys* Vol. 2 (eds Hardisty, M. W. & Potter, I. C.) 373–404 (Academic, London, 1972).
- Janvier, P. *Early Vertebrates* (Clarendon, Oxford, 1996).
- Shu, D.-G. *et al.* Lower Cambrian vertebrates from south China. *Nature* **402**, 42–46 (1999).
- Mallatt, J. & Sullivan, J. 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. *Mol. Biol. Evol.* **15**, 1706–1718 (1998).
- Kuraku, S. *et al.* Monophyly of lampreys and hagfishes supported by nuclear DNA-coded genes. *J. Mol. Evol.* **49**, 729–735 (1999).
- Zhang, X.-L. *et al.* New sites of Chengjiang fossils: crucial windows on the Cambrian explosion. *J. Geol. Soc. Lond.* **158**, 211–218 (2001).
- Janvier, P. Catching the first fish. *Nature* **402**, 21–22 (1999).
- Shimeld, S. M. & Holland, P. W. H. Vertebrate innovations. *Proc. Natl Acad. Sci. USA* **97**, 4449–4452 (2000).
- Holland, H. D. & Chen, J.-y. Origin and early evolution of the vertebrates: new insights from advances in molecular biology, anatomy, and palaeontology. *BioEssays* **23**, 142–151 (2001).
- Cole, W. C. & Youson, J. H. Morphology of the pineal complex of the anadromous sea lamprey, *Petromyzon marinus* L. *Am. J. Anat.* **165**, 131–163 (1982).
- Bardack, D. & Zangerl, R. *The Biology of Lampreys* Vol. 1 (eds Hardisty, M. W. & Potter, I. C.) 67–84 (Academic, London, 1971).
- Gabbott, S. E., Aldridge, R. J. & Theron, J. N. A giant conodont with preserved muscle tissue from the Upper Ordovician of South Africa. *Nature* **374**, 800–803 (1995).
- Gagnier, P.-Y. *Sacabambaspis janvieri*, vertébré Ordovicien de Bolivie: I. Analyse morphologique. *Ann. Paléontol.* **79**, 19–51 (1993).
- Gagnier, P.-Y. *Sacabambaspis janvieri*, vertébré Ordovicien de Bolivie 2: Analyse phylogénétique. *Ann. Paléontol.* **79**, 119–166 (1993).
- Ritchie, A. New evidence on *Jamoytius kerwoodi* White, an important ostracoderm from the Silurian of Lanarkshire, Scotland. *Palaeontology* **11**, 21–39 (1968).
- Lacalli, T. C. Frontal eye circuitry, rostral sensory pathways and brain organization in amphioxus larvae: evidence from 3D reconstructions. *Phil. Trans. R. Soc. Lond. B* **351**, 243–263 (1996).
- Lacalli, T. C. New perspectives on the evolution of protochordate sensory and locomotor systems, and the origin of brains and heads. *Phil. Trans. R. Soc. Lond. B* **356**, 1565–1572 (2001).
- Fritsch, B. Evolution of the vestibulo-ocular system. *Otolaryngology Head Neck Surg.* **119**, 182–192 (1998).
- Kuratani, S. *et al.* Embryology of the lamprey and evolution of the vertebrate jaw: insights from molecular and developmental perspectives. *Phil. Trans. R. Soc. Lond. B* **356**, 1615–1632 (2001).
- Cohn, M. J. Lamprey *Hox* genes and the origin of jaws. *Nature* **416**, 386–387 (2002).
- Shigetani, Y. *et al.* Heterotopic shifts of epithelial–mesenchymal interactions in vertebrate jaw evolution. *Science* **296**, 1316–1319 (2002).
- Shu, D.-G. *et al.* Primitive deuterostomes from the Chengjiang Lagerstätte (Lower Cambrian, China). *Nature* **414**, 419–424 (2001).
- Shu, D.-G. *et al.* An early Cambrian tunicate from China. *Nature* **411**, 472–473 (2001).
- Shu, D.-G., Conway Morris, S. & Zhang, X.-L. A *Pikaia*-like chordate from the Lower Cambrian of China. *Nature* **384**, 157–158 (1996).
- Janvier, P. The dawn of the vertebrates: characters versus common ascent in the rise of current vertebrate phylogenies. *Palaeontology* **39**, 259–287 (1996).
- Farris, J. S. HENNIG86, version 1.5. Program and user's manual (published by the author, Port Jefferson Station, New York, 1988).
- Courrol Ramos, T. TREE GARDENER, version 1.0. (<http://www.icn.unal.edu.co/extensio/servicio/servicio.html>) (1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This work was supported by the Ministry of Sciences and Technology of China, the National Science Foundation of China, the Ministry of Education of China, the National Geographic Society (USA), The Royal Society, and St John's College, Cambridge. We thank K. Kardong and B. J. Swalla. L. Guo, X. Cheng, M. Cheng and S. Last are thanked for technical assistance, and Y. Ji and H. Guo for fieldwork help.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to D.-G.S. (e-mail: dgidshu@nwnu.edu.cn).

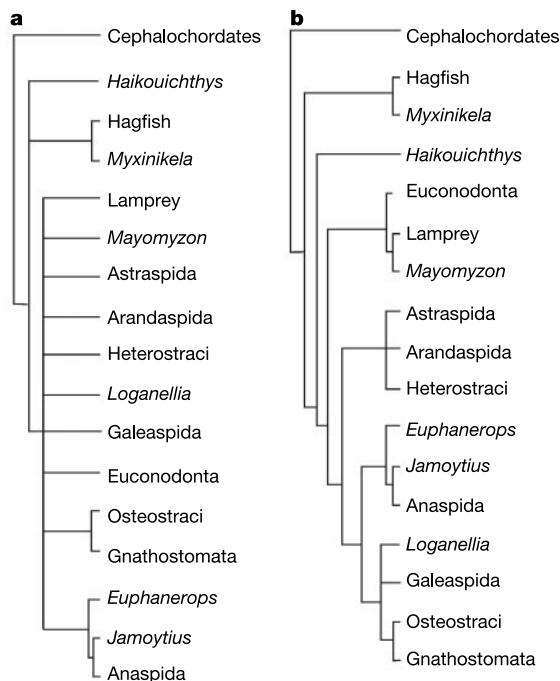


Figure 2 Phylogenetic analysis. **a**, Strict consensus of 23 equally parsimonious trees (length 177 steps; consistency index 0.64, retention index 0.64). *Haikouichthys* appears here in a trichotomy with hagfishes and all other vertebrates (that is, one possibility is that it is a stem craniate), but this is largely because of the inferred presence of metamerically gonads. **b**, Strict consensus of four equally parsimonious trees (length 175, consistency index 0.64, retention index 0.64) obtained when the character 'metameric gonads' (114) is inactivated. *Haikouichthys* appears here as the sister-group to all other vertebrates except hagfishes, like *Mylokunmingia* in the analysis of ref. 5.

Effects of household dynamics on resource consumption and biodiversity

Jianguo Liu*, Gretchen C. Daily†, Paul R. Ehrlich† & Gary W. Luck‡

* Department of Fisheries and Wildlife, Michigan State University, E. Lansing, Michigan 48824, USA

† Center for Conservation Biology, Stanford University, Stanford, California 94305, USA

Human population size and growth rate are often considered important drivers of biodiversity loss^{1–6}, whereas household dynamics are usually neglected. Aggregate demographic statistics may mask substantial changes in the size and number of households, and their effects on biodiversity. Household dynamics influence per capita consumption^{7,8} and thus biodiversity through, for example, consumption of wood for fuel⁹, habitat alteration for home building and associated activities^{10–12}, and greenhouse gas emissions¹³. Here we report that growth in household numbers globally, and particularly in countries with biodiversity hotspots (areas rich in endemic species and threatened by human activities¹⁴), was more rapid than aggregate population growth between 1985 and 2000. Even when population size declined, the number of households increased substantially. Had the average household size (that is, the number of occupants) remained static, there would have been 155 million fewer households in hotspot countries in 2000. Reduction in average household size alone will add a projected 233 million additional households to hotspot countries during the period 2000–15. Rapid increase in household numbers, often manifested as urban sprawl, and resultant higher per capita resource consumption in smaller households^{15–19} pose serious challenges to biodiversity conservation.

As a first step towards quantifying the effects of household dynamics on biodiversity, we compared the rates of change in human population size and the number of households in 76 hotspot and 65 non-hotspot countries. We also investigated the sources of growth in household numbers, comparing the relative contributions of changes in aggregate population size and household size. Finally, in six representative hotspot areas, we estimated the relative contributions of changes in population size and household size to the growth in the number of households (see Methods).

In hotspot countries, the annual rate of growth in the number of households (3.1%) was substantially higher than the population growth rate (1.8%) between 1985 and 2000 (Fig. 1a). Over 80% of

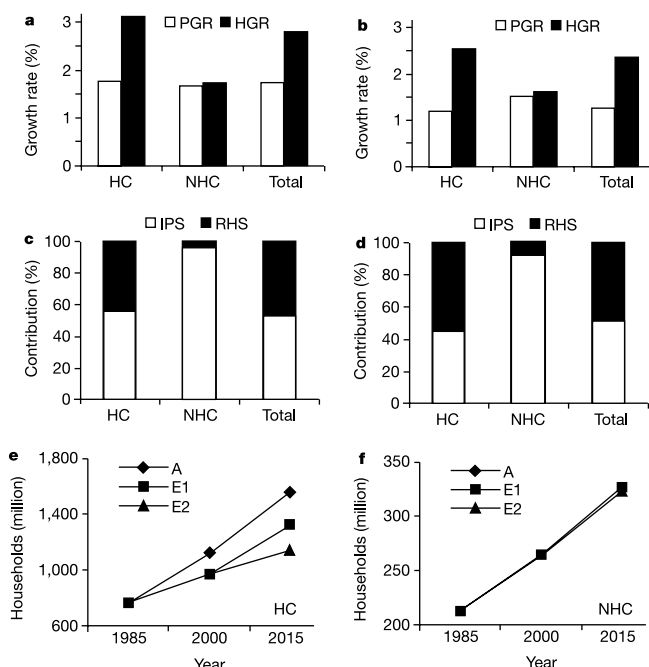


Figure 1 Household dynamics in 76 hotspot countries (HC) and 65 non-hotspot countries (NHC). **a**, Annual growth rates of population size (PGR) and household number (HGR) from 1985–2000. **b**, Projected PGR and HGR from 2000–15. **c**, Contributions of reduction in household size (RHS) and increase in population size (IPS) to household numbers during 1985–2000. **d**, Projected contributions of RHS and IPS from 2000–15. In **e** (HC) and **f** (NHC), 'A' indicates the actual household number; E1 is the estimated household number assuming that the average household size in 2000 was at the 1985 level and that the average household size in 2015 remains at the level in 2000; E2 assumes that average household sizes in both 2015 and 2000 were at the 1985 level.

hotspot countries showed this pattern (Table 1). In contrast, population and household growth rates in non-hotspot countries were roughly equivalent (1.7%) (Fig. 1a). The divergence in population and household growth rates is expected to become more pronounced over the next 15 years (Fig. 1b and Table 1), suggesting that it is crucial to consider growth in the number of households when assessing threats to biodiversity.

The growth in household number resulted directly from a simultaneous increase in population size and reduction in average household size. In 1985, average household size was larger by 1.0 persons in hotspot countries (4.7) than in non-hotspot countries (3.7). This difference was reduced to 0.3 persons in 2000. By 2015,

Table 1 Comparisons between rates of growth in household number and population size in 76 hotspot and 65 non-hotspot countries

Time period (years)	Relationship between <i>hnn</i> and <i>pop</i>	Hotspot countries		Non-hotspot countries		Total	
		Number of countries	Per cent	Number of countries	Per cent	Number of countries	Per cent
1985–2000	<i>hnn</i> (+) > <i>pop</i> (+)*	63	82.9	42	64.6	105	74.5
	<i>hnn</i> (+) < <i>pop</i> (+) [†] ‡	11	14.5	19	29.2	30	21.3
	<i>hnn</i> (+) & <i>pop</i> (–)‡	1	1.3	3	4.6	4	2.8
	<i>hnn</i> (–) & <i>pop</i> (+)§	1	1.3	1	1.6	2	1.4
2000–15	<i>hnn</i> (+) > <i>pop</i> (+)	67	88.2	46	70.8	113	80.1
	<i>hnn</i> (+) < <i>pop</i> (+)	5	6.6	12	18.5	17	12.1
	<i>hnn</i> (+) & <i>pop</i> (–)	4	5.2	6	9.2	10	7.1
	<i>hnn</i> (–) & <i>pop</i> (+)	0	0.0	1	1.5	1	0.7

Abbreviations: *hnn*, rate of growth in household number; *pop*, rate of growth in population size. The frequency of hotspot and non-hotspot countries occurring in the categories *hnn*(+) > *pop*(+) and *hnn*(+) < *pop*(+) differed at $P = 0.0365$ for 1985–2000 and $P = 0.0341$ for 2000–15.

*Both *hnn* and *pop* are positive, and the former is greater than the latter.

†Both *hnn* and *pop* are positive and the former is smaller than the latter.

‡*hnn* is positive whereas *pop* is negative.

§*hnn* is negative whereas *pop* is positive.

the average household size in hotspot countries (3.4) is expected to be 0.2 persons smaller than that of non-hotspot countries (3.6). Per cent contribution of reduced household size to growth in the number of households was about 12 times higher in hotspot countries than in non-hotspot countries in 1985–2000 (Fig. 1c), and it is estimated to be about 7 times higher in hotspot countries than in non-hotspot countries in the period 2000–15 (Fig. 1d). Furthermore, this contribution is projected to increase from 43% (hotspot countries) and 3% (non-hotspot countries) in 1985–2000 (Fig. 1c) to 54% and 7%, respectively, in 2000–15 (Fig. 1d). Most countries containing hotspots have relatively low population growth rates, and the primary demographic pressure on their biodiversity will come from urban sprawl and other impacts associated with increased household numbers.

Had the average household size remained at the 1985 level, there would have been 155 million fewer households in hotspot countries in 2000 (Fig. 1e). By 2015, 233 million more households are likely to be added to hotspot countries as a result of continued reduction in average household size alone. If the average household size in 2015 were the same as in 1985, there would be 415 million fewer households in hotspot countries (Fig. 1e); in non-hotspot countries, there would be 4 and 7 million fewer households in 2000 and 2015, respectively (Fig. 1f). In four hotspot countries (Italy, Portugal, Spain and Greece) over the period 2000–15, contributions of reduction in average household size to growth in the number of households are expected to be roughly 120–190% (equivalent to 0.4–2.4 million additional households in each country) even when their corresponding population sizes are projected to decline at an annual rate of 0.1–0.3%.

In the six representative hotspot areas listed in Table 2, reduction in average household size contributed approximately 30–73% to the growth in the number of households over the periods of 10 and 40 years. Annual rates of population growth in these six areas ranged from 0.5% to 7.0%, whereas annual rates of growth in household numbers were much higher (1.7–10.0%) owing to a decline in average household size (0.7–1.2% per year). By 1991, Italy had added almost 6 million households as a result of reduced average household size alone since 1951 (Table 2). In Brazil, over 4.6 million

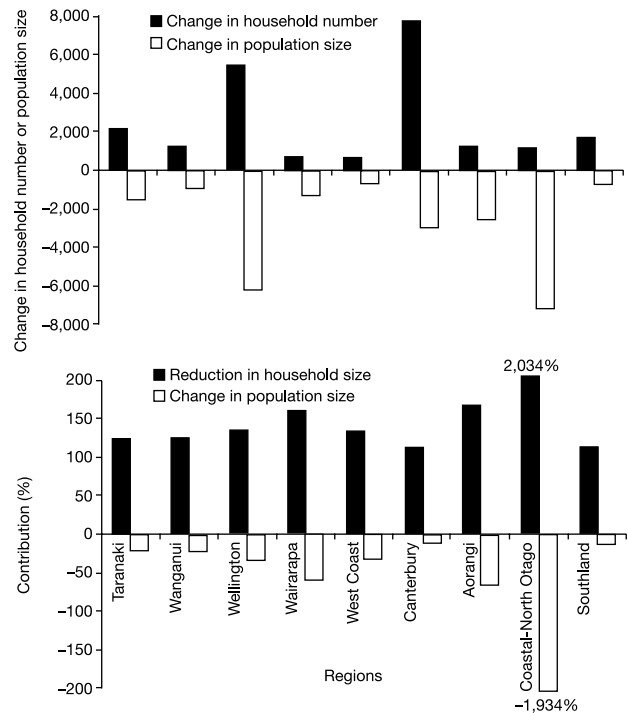


Figure 2 Changes in household numbers and population size (top) as well as their contributions to growth in household number (bottom) in nine regions of New Zealand.

households were created between 1991 and 2000 owing to a reduction in average household size. Had the average household size stayed at the 1970 level, Indian River County (United States) would have had 11,103 fewer households in 2000. These extra households are among the factors that have made Indian River County one of the endangered species hotspots in the United States (areas with the largest numbers of federally listed endangered and

Table 2 Dynamics of human populations and households in six biodiversity hotspot areas

	New Zealand	Italy	Brazil	Indian River County, USA	Island of Rodrigues, Mauritius	Wolong Nature Reserve, China
Relationship with biodiversity hotspots	Is identical to 'New Zealand'	Contains a portion of 'Mediterranean Basin'	Contains a portion of 'Brazilian Cerrado' and 'Atlantic Forest Region'	Contains a portion of 'Caribbean'	Is part of 'Madagascar and Indian Ocean Islands'	Is part of 'Mountains of South-Central China'
Time period						
T_0	1976	1951	1991	1970	1983	1975
T_1	1991	1991	2000	2000	2000	1998
Population size						
T_0	2,975,846	47,516,000	145,605,330	35,743	32,925	2,560
T_1	3,237,915	56,322,185	168,370,893	110,558	35,769	4,320
Average annual growth rate of population size (%)	0.6	0.5	1.7	7.0	0.5	3.0
Average household size						
T_0	3.22	4.02	4.95	2.90	4.95	6.08
T_1	2.75	2.83	3.76	2.25	4.13	4.95
Average annual rate of growth of household size (%)	-1.0	-0.7	-1.2	-0.8	-1.0	-1.1
Number of households						
T_0	923,257	11,814,402	34,734,715	12,325	6,652	421
T_1	1,177,665	19,909,003	44,795,101	49,137	8,651	942
Average annual rate of growth in the number of households (%)	1.8	1.7	3.2	10.0	1.8	5.4
Extra number of households due to reduction in average household size	172,101	5,905,027	4,629,573	11,103	1,424	231
Contribution of reduction in average household size to growth in number of households (%)	67.7	73.0	46.0	30.0	71.3	44.4

*The remainder of the contribution came from population growth.

threatened species)²⁰. Reduction in average household size was an important factor causing the increase of household number and thus rise in the amount of fuel wood consumed in Wolong Nature Reserve (China), which contributed to increased deforestation and loss and fragmentation of habitat for giant pandas²¹. In Wolong, there would have been approximately 230 fewer households in 1998 if average household size had been kept at the 1975 level (Table 2).

Even in regions where population size decreased, the number of households still increased substantially owing to a reduction in average household size. During the period 1976–81, population size in nine government regions of New Zealand declined by about 640–7,200 people, but the number of households rose by approximately 560–7,650 per region (Fig. 2, top). Contributions of reduction in average household size to the increase in the number of households ranged from roughly 112–2,034%, compared with –1,934% to –12% resulting from the change in population size (Fig. 2, bottom). In the other 13 regions of New Zealand where the population increased, reduction in average household size contributed about 29–82% to the growth in the number of households, whereas population growth accounted for approximately 18–71%. Considering all 22 New Zealand government regions, contributions of reduction in average household size and change in population size to growth in household number were about 83% and 17%, respectively.

Reduction in average household size takes a double toll on resource use and biodiversity. First, more households mean more housing units, thus generally increasing the amount of land and materials (for example, wood, concrete and steel) needed for housing construction (see Supplementary Results). Second, smaller households have lower efficiency of resource use per capita (Fig. 3) because goods and services are shared by more people in larger households^{16–18} (see Supplementary Results). Proximate causes of a reduction in household size include lower fertility rates, higher per capita income, higher divorce rates, ageing populations, and a decline in the frequency of multi-generational families living together (see refs 22–24). Some of these factors can affect both population growth and household growth. Although lower fertility rates may reduce population growth and household numbers, the resulting potential reduction in resource consumption may be offset by higher per capita consumption in smaller households. Thus, declining fertility rates are necessary but not sufficient to ensure reduced anthropogenic pressure on the environment and natural landscapes. Our study suggests that biodiversity conservation is faced with much larger challenges than previously thought, because reduction in household size leads to higher per capita resource consumption and a rapid increase in the number of households, even when population size declines. This trend is most prevalent in hotspot countries where it may severely limit efforts to conserve biodiversity, thus degrading the ecosystem services²⁵ that biodiversity delivers to humanity. □

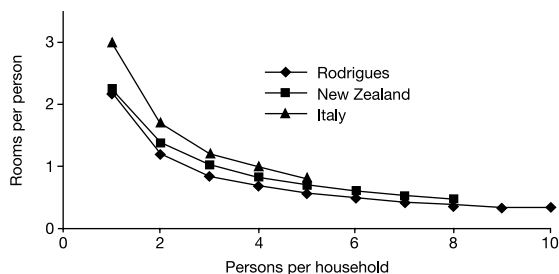


Figure 3 Rooms per capita versus household size in New Zealand, Italy and Rodrigues of Mauritius.

Methods

Data on population sizes and household numbers for 141 countries over the period 1985–2015 are from the United Nations²⁶. The hotspot countries were identified according to Conservation International (<http://www.biodiversityhotspots.org/xp/Hotspots/>) and ref. 27, confirmed by N. Myers and P. Langhammer (personal communication), and are listed in Supplementary Table 1. The six areas used in our detailed analysis were chosen from six major regions (Africa, Asia, Europe, North America, Oceania and South America) on the basis of data availability, representation and policy implications (see Supplementary Methods).

Rates of growth and per cent contributions in hotspot and non-hotspot countries (Fig. 1) were weighted by population size and number of households in each country because of great variations among countries. To compare the rates of growth in household number (*hln*) and population size (*pop*) (Table 1), we determined the number and percentage of hotspot and non-hotspot countries with $hln(+) > pop(+)$ and $hln(+) < pop(+)$ (where + indicates positive rates of growth). Using Fisher's exact test, we tested for differences in the frequency of occurrence of hotspot and non-hotspot countries in these categories, for 1985–2000 and 2000–15, respectively.

Changes in the number of households are affected directly by changes in population size and household size. We calculated the per cent contribution of change in average household size (*chs*) to the change in the number of households (H_{chs}) as the total contribution (100%) minus the contribution due to population growth alone:

$$H_{chs} = [H_1 - H_0 - H_P] / [H_1 - H_0] \times 100\% = 100\% - H_P / [H_1 - H_0] \times 100\% \quad (1)$$

where H_0 and H_1 are the number of households at time 0 and 1, respectively. H_P is the growth in the number of households (with the same average household size at time 0 (S_0)) due to population growth or the difference in populations ($P_1 - P_0$) at times 1 and 0: $H_P = (P_1 - P_0) / S_0$. The average household size was fixed at its value for time 0 (S_0) (when H_P was computed) to see how many fewer households would exist at time 1 if the average household size remained static. As in standard decomposition techniques²⁸, taking a simple average of household sizes at times 0 and 1 would be another way to calculate H_P . However, our choice of S_0 yields similar results and is the most appropriate method to address our question.

The contribution of change in population size (*cps*) to growth in the number of households (H_{cps}) is the total contribution minus the contribution owing to change in household size:

$$H_{cps} = 100\% - H_{chs} = 100\% - (100\% - H_P / [H_1 - H_0] \times 100\%) = H_P / [H_1 - H_0] \times 100\% \quad (2)$$

Received 6 September; accepted 12 December 2002; doi:10.1038/nature01359.

Published online 12 January 2003.

- Holdren, J. P. & Ehrlich, P. R. Human population and the global environment. *Am. Sci.* **62**, 282–292 (1974).
- Wilson, E. O. *Biodiversity* (National Academy of Science, Washington DC, 1988).
- Pimm, S. L., Russell, G. J., Gittleman, H. L. & Brooks, T. M. The future of biodiversity. *Science* **269**, 347–350 (1995).
- Thompson, K. & Jones, A. Human population density and prediction of local plant extinction in Britain. *Conserv. Biol.* **13**, 185–189 (1999).
- Cincotta, R. P., Wisniewski, J. & Engelman, R. Human population in the biodiversity hotspots. *Nature* **404**, 990–992 (2000).
- Dompka, V. *Human Population, Biodiversity and Protected Areas: Science and Policy Issues* (American Association for the Advancement of Science, Washington DC, 1996).
- Kaul, S. & Liu, Q. Rural household energy use in China. *Energy* **17**, 405–411 (1992).
- Sandiford, P., Gorter, A. C., Orozco, J. G. & Pauw, J. P. Determinants of domestic water-use in rural Nicaragua. *J. Trop. Med. Hyg.* **93**, 383–389 (1990).
- Gardner-Outlaw, T. & Engelman, R. *Forest Futures: Population, Consumption and Wood Resources* (Population Action International, Washington DC, 1999).
- Friesen, L. E., Eagles, P. F. J. & MacKay, R. J. Effects of residential development on forest dwelling neotropical migrant songbirds. *Conserv. Biol.* **9**, 1408–1414 (1995).
- Nilon, C. H., Long, C. N. & Zipperer, W. C. Effects of wildland development on forest bird communities. *Landscape Urban Plan.* **32**, 81–92 (1995).
- Kluza, D. A., Griffin, C. R. & DeGraaf, R. M. Housing developments in rural New England: effects on forest birds. *Anim. Conserv.* **3**, 15–26 (2000).
- MacKellar, F. L., Lutz, W., Prinz, C. & Goujon, A. Population, households, and CO₂ emissions. *Pop. Dev. Rev.* **21**, 849–865 (1995).
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
- An, L. *et al.* Simulating demographic and socioeconomic processes on household level and implications for giant panda habitats. *Ecol. Mod.* **140**, 31–50 (2001).
- Yousif, H. M. Population, biomass and the environment in central Sudan. *Int. J. Sust. Dev. World Ecol.* **2**, 54–69 (1995).
- Lenzen, M. & Murray, S. A. A modified ecological footprint method and its application to Australia. *Ecol. Econ.* **37**, 229–255 (2001).
- Ironmonger, D. S., Aitken, C. K. & Erbas, B. Economies of scale in energy use in adult-only households. *Energy Econ.* **17**, 301–310 (1995).
- Durrenberger, G., Patzel, N. & Hartmann, C. Household energy consumption in Switzerland. *Int. J. Environ. Pollut.* **15**, 159–170 (2001).
- Rutledge, J., Lepczyk, C., Xie, J. & Liu, J. Spatial and temporal dynamics of endangered species hotspots in the United States. *Conserv. Biol.* **15**, 475–487 (2001).
- Liu, J. *et al.* Ecological degradation in protected areas: The case of Wolong Nature Reserve for giant pandas. *Science* **292**, 98–101 (2001).
- Michael, R. T., Fuchs, V. R. & Scott, S. R. Changes in the propensity to live alone: 1950–1976. *Demography* **17**, 39–53 (1980).

23. Bongaarts, J. *The End of the Fertility Transition in the Developed World* (The Population Council, New York, 2001).
24. Kinsella, K. & Velkoff, V. A. *An Aging World: 2001* (US Government Printing Office, Washington DC, 2001).
25. Daily, G. C. (ed.) *Nature's Services: Societal Dependence on Natural Ecosystems* (Island, Washington DC, 1997).
26. United Nations Centre for Human Settlements (Habitat) *Cities in a Globalizing World: Global Report on Human Settlements 2001* (Earthscan, London, 2001).
27. Cincotta, R. P. & Engelman, R. *Nature's Place: Human Population and the Future of Biological Diversity* (Population Action International, Washington DC, 2000).
28. Kitagawa, E. M. Components of a difference between two rates. *J. Am. Stat. Assoc.* **50**, 1168–1194 (1955).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank J. Eagle, W. Falcon, M. Feldman, N. Keilman, H. Mooney, R. Naylor, S. Pimm, K. Seto and S. Tuljapourkar for their constructive comments on earlier drafts; P. Langhammer and N. Myers for providing lists of hotspot countries; E. Laurent for technical assistance in producing figures; J. Baca, R. Cincotta, W. Lutz and A. McMillan for providing some references; G. Clarke for providing the housing data of India River County, Florida; and W. W. Taylor and Q. Wang for logistical and moral support. Funding for this project was provided to J.L. by the National Science Foundation (CAREER Award and Biocomplexity in the Environment Program) and the National Institute of Child Health and Human Development.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.L. (e-mail: jlui@panda.msu.edu).

The genetic basis of family conflict resolution in mice

Reinmar Hager & Rufus A. Johnstone

Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Asymmetries in the costs and benefits of parental investment for mothers, fathers and offspring result in family conflict over the production and provisioning of young^{1–3}. In species where females provide most resources before and after birth, the resolution of this conflict may be influenced by genes expressed in mothers and by maternally and paternally inherited genes expressed in offspring^{4,5}. Here we disentangle these effects by means of reciprocal mating and cross-fostering of litters between two strains of mice that differ with respect to the typical resolution of family conflict. We find that differences in litter size between these two strains are determined by paternal genotype, whereas differences in provisioning are under maternal control, showing that there is antagonistic coadaptation of maternal and paternal effects on distinct life-history traits. Maternal provisioning is also influenced by the type of foster offspring. Contradictory to theoretical expectations, however, we find no evidence for a negative correlation across strains between maternal provisioning and offspring demand. Instead, we show that there is positive coadaptation such that offspring obtain more resources from foster mothers of the same strain as their natural mother, irrespective of their father's strain.

Offspring are typically selected to demand more resources from parents than parents are selected to provide^{1,2,6}. In addition to this parent–offspring conflict over provisioning, conflicts will often arise between mothers and fathers^{7,8}. Each parent favours greater investment by the other than is in the other's best interest⁹. Even where males provide little or no direct care, this sexual conflict can manifest itself in parent-of-origin-specific effects on solicitation

behaviour in offspring that affect resource transfer from mothers, for example, through genomic imprinting^{5,10,11}. Biologists have begun to investigate the genetic basis of these family conflicts and their resolution; previous studies have focused either on sexual conflict between paternally and maternally inherited genes¹⁰, or on coadaptation of maternal and offspring genes^{4,12}. Here we report the results of a study that has assessed simultaneously the influence on maternal investment of genes expressed in mothers, and maternally and paternally inherited genes expressed in offspring.

We conducted a series of crosses using mice of two strains, CBA and C57/B6 (hereafter called B6), that differ in the resolution of family conflict. CBA litters typically contain more pups than B6 litters (Fig. 1), but the individual birth weight of CBA pups is typically less than that of B6 pups (Fig. 2). To determine the contribution of maternal genes and of maternally and paternally inherited offspring genes to differences between the strains in maternal investment, we carried out crosses that included pure strain CBA matings, pure strain B6 matings, and both possible reciprocal crosses (male CBA with female B6, and female CBA with male B6). The resulting 117 litters were all cross-fostered among (unrelated) females, yielding foster families that included all eight possible combinations of foster mother's strain, natural mother's strain and father's strain.

We recorded the size of each litter and measured maternal provisioning 6 d after birth, a stage at which pups gain weight only by suckling. After simulated departure of the foster mother for

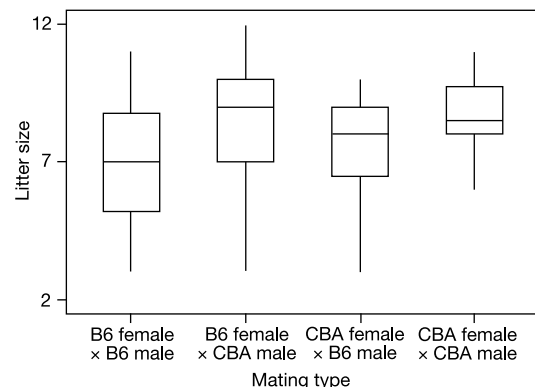


Figure 1 Litter sizes produced by all four possible types of mating (B6 female × B6 male, B6 female × CBA male, CBA female × B6 male and CBA female × CBA male).

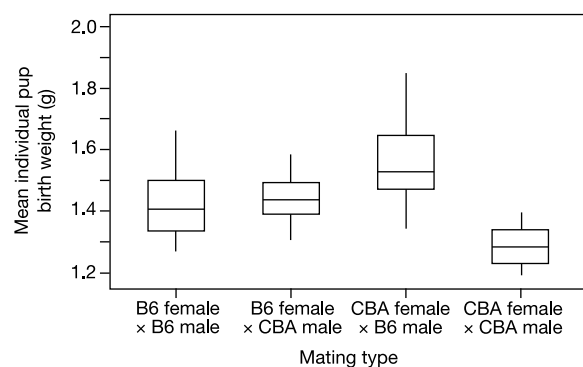


Figure 2 Mean individual birth weight of pups in litters produced by all four possible types of mating (B6 female × B6 male, B6 female × CBA male, CBA female × B6 male and CBA female × CBA male). Note that each litter contributes a single mean value of pup birth weight to the data set.

Table 1 GLM analysis of litter size

	d.f.	F	P
Full model			
Father's strain	1	14.04	0.000
Body weight of mother	1	6.81	0.010
Mother's strain	1	1.24	0.267
Father's strain × mother's strain	1	0.16	0.691
Minimal model	Coefficient	SE coefficient	
Constant	2.48	2.01	
Father's strain	- 0.77	0.21	
Body weight of mother	0.26	0.01	

Percentage variance accounted for 15.72 Total d.f. 117

d.f., degrees of freedom; SE, standard error.

4h, provisioning was measured in 35 litters by the loss in foster mother's body weight during the 2 h after she rejoined her adoptive pups (females did not consume any food over this interval of time and started nursing soon after they rejoined the pups). General linear model (GLM) analysis was used to assess the influence of natural mother's strain, father's strain and interactions between the two on litter size (including natural mother's body weight as a covariate), and to assess the influence of foster mother's strain, natural mother's strain, father's strain and all interactions between these factors on foster mothers' loss of body weight (including foster mother's body weight, strain of foster mother's mate, foster litter size and sex ratio, and natural mother's body weight in the analysis as covariates).

Results from the GLM analysis of litter size are summarized in Table 1. Natural mother's body weight was found to have a significant effect on litter size (GLM; $F_{1,113} = 6.81$; $P = 0.010$), with larger females giving birth to bigger litters (although little of the variation in litter size could be explained in this way; $r^2 = 4.9\%$). Apart from the effect of maternal body weight, father's strain type was the only factor found to have a significant influence on litter size (GLM; $F_{1,113} = 14.04$, $P < 0.001$), with CBA males siring larger litters than B6 males (Fig. 1); (natural) mother's strain type did not influence litter size (and we also note that there was no significant difference in body weight between mothers of the two strains).

Results from the GLM analysis of maternal provisioning are summarized in Table 2. A foster mother's strain had a significant effect on her weight loss after she rejoined her adoptive pups (GLM; $F_{1,24} = 31.47$; $P < 0.001$), with CBA females providing fewer resources than B6 females (Fig. 3). Provisioning was also influenced by foster litter type. But natural mother's strain, father's strain or strain of foster mother's mate did not have a major effect on the weight loss of a litter's foster mother. Instead, foster mother weight loss was significantly affected by an interaction between foster mother's strain and natural mother's strain (GLM; $F_{1,24} = 12.07$; $P = 0.004$), with females that cared for litters produced by a mother of their own strain losing more weight (Fig. 3); there was no such interaction between foster mother's strain and father's strain.

The above results provide support for the idea of antagonistic co-evolution between the sexes, in that a positive paternal influence on maternal investment (siring of larger litters by CBA males) is counterbalanced by a negative maternal influence (reduced provisioning by CBA females). Uniquely, however, these effects involve distinct life-history traits: litter size and postnatal provisioning (by contrast, previous studies¹⁰ have examined opposing maternal and paternal influences on the same trait, prenatal provisioning, mediated by the paternally expressed gene *Peg3*). Our results show that the trade-off between offspring size and number is influenced by males, whereas females adjust postnatal provisioning to balance the negative effects of increased litter size on their fitness.

The CBA strain is characterized by a male trait, increased litter size, that serves to elicit greater investment from mothers. Although larger litter size does not entail greater litter weight at birth (because individual pup birth weight is negatively related to the number of

Table 2 GLM analysis of foster mother's weight loss in relation to her strain* and the strain of her adoptive litter's mother and father

	d.f.	F	P
Full model			
Foster mother's strain	1	31.47	0.000
Foster mother's strain × mother's strain†	1	12.07	0.004
Foster mother's strain × father's strain†	1	2.32	0.139
Litter size	1	1.98	0.169
Foster mother's mate's strain†	1	1.77	0.193
Body weight of litter's natural mother	1	1.06	0.311
Father's strain × mother's strain†	1	0.34	0.564
Father's strain	1	0.11	0.745
Arcsine proportion of males in litter	1	0.09	0.763
Body weight of foster mother	1	0.03	0.855
Mother's strain	1	0.00	0.993
Minimal model	Coefficient	SE coefficient	
Constant	1.28	0.10	
Foster mother's strain	0.58	0.10	
Foster mother's strain × mother's strain†	0.36	0.10	

Percentage variance accounted for 56.96 Total d.f. 35

*Weight loss was measured after the return of adoptive pups.

†Mother's strain and father's strain refer to the strain of the natural mother and father of the litter raised by the foster mother in question.

young), there is a significant positive impact of pup number on the total weight gain of the litter from days 1 to 5 (GLM; $F_{1,30} = 25.14$, $P < 0.001$). Larger litters thus entail greater postnatal investment by a female, confirming results of previous studies¹³.

Greater investment by a female is beneficial for her mate, assuming that he does not mate with her again, because he will not suffer from any reduction in her reproductive success owing to investment in the current litter. But it is likely (under natural conditions) to prove costly for the female^{13,14}. Consequently, we might expect maternal counteradaptation. Our finding that CBA mothers provision less after rejoining pups supports this prediction. The basis of the paternal influence on litter size remains to be elucidated but, because females can resorb embryos¹⁵, we suggest that expression of paternally expressed genes in developing young may influence the likelihood of resorption.

In regard to the influence of young on provisioning, models of maternal-offspring coadaptation suggest that coadaptation of maternal and offspring traits should lead to a negative correlation between the effects of genes expressed in offspring and mothers^{16,17}. Our results do not support this prediction. B6 mothers, who are better providers than CBA females, do not produce offspring that are weaker elicitors. The absence of any negative correlation is to be expected, however, given the antagonistic coadaptation between the sexes described above. Because reduced provisioning by CBA mothers is counterbalanced by the larger size of litters sired by CBA males, selection for stronger elicitation by the offspring of such mothers will be weak or absent. This finding highlights the importance of examining different aspects of family conflict in

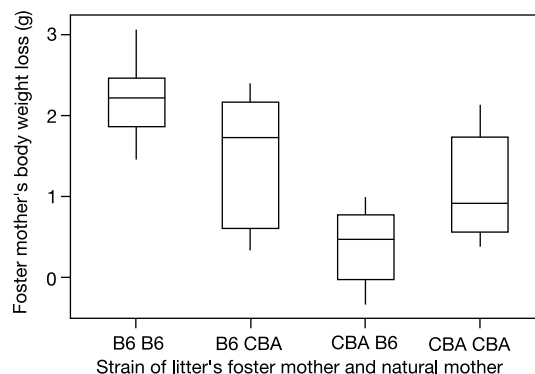


Figure 3 Foster mother's loss of body weight during the 2 h after rejoining her adoptive litter, depending on her strain and that of the natural mother of her foster pups.

conjunction, rather than studying either parent–offspring or male–female co-evolution in isolation.

We also find evidence for more complex, positive coadaptation of maternal and offspring effects on resource transfer, in that offspring are more successful in obtaining food from a foster mother of the same strain as their natural mother (regardless of their father's strain). This could be attributed to differences between the strains in pup behaviour, with young better adapted to elicit provisioning from mothers of their own strain. Alternatively, it might reflect selection for kin discrimination on the part of mothers. In mice, young are often nursed in communal nests with females caring for alien offspring¹⁸. Under these circumstances, it would be beneficial for a female to recognize and favour her own young. Only the maternally inherited component of the offspring genotype is relevant to this discrimination task, potentially explaining the absence of any paternal influence on pup weight gain in our study. We note also that advertisement by pups of their maternal origin might prove beneficial under the conditions described.

In summary, we find strong evidence for coadaptation of maternal genes and both maternally and paternally inherited genes in mice. The close interaction between embryo and mother *in utero*, combined with asymmetrical costs and benefits of maternal investment for offspring, mother and father in mammals, results in a complex array of family conflicts. Because paternal genes can affect maternal investment (for example, by influencing litter size or prenatal growth) we suggest that paternal effects should be considered alongside maternal effects even when there is only maternal care of young. In particular, the effects of imprinted genes should be incorporated in theories of maternal coadaptation. Future studies should not restrict their focus to parent–offspring or male–female conflict in isolation. □

Received 8 July; accepted 8 October 2002; doi:10.1038/nature01239.

1. Trivers, R. L. Parent–offspring conflict. *Am. Zool.* **14**, 249–264 (1974).
2. Godfray, H. C. J. Evolutionary theory of parent–offspring conflict. *Nature* **376**, 133–138 (1995).
3. Parker, G. A., Royle, N. J. & Hartley, I. R. Intrafamilial conflict and parental investment: a synthesis. *Phil. Trans. R. Soc. Lond. B* **357**, 295–307 (2002).
4. Agrawal, A. F., Brodie, E. D. & Brown, J. Parent–offspring coadaptation and the dual genetic control of maternal care. *Science* **292**, 1710–1712 (2001).
5. Haig, D. Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proc. R. Soc. Lond. B* **264**, 1657–1661 (1997).
6. Trivers, R. L. *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 139–179 (Aldine, Chicago, 1972).
7. MacNair, M. R. & Parker, G. A. Models of parent–offspring conflict II. Promiscuity. *Anim. Behav.* **26**, 111–122 (1978).
8. Lessells, C. M. Parentally biased favouritism: why should parents specialise in caring for different offspring. *Phil. Trans. R. Soc. Lond. B* **357**, 381–403 (2002).
9. Houston, A. I. & Davies, N. B. *Behavioural Ecology* (eds Sibley, R. M. & Smith, R. H.) 471–487 (Blackwell Scientific, Oxford, 1985).
10. Li, L. L. *et al.* Regulation of maternal behaviour and offspring growth by paternally expressed Peg3. *Science* **284**, 330–333 (1999).
11. Keverne, E. B. Genomic imprinting, maternal care, and brain evolution. *Horm. Behav.* **40**, 146–155 (2001).
12. Kölliker, M. & Richner, H. Parent–offspring conflict and the genetics of offspring solicitation and parental response. *Anim. Behav.* **62**, 395–407 (2001).
13. Fuchs, S. Optimality of parental investment: The influence of nursing on reproductive success of mother and female young house mice. *Behav. Ecol. Sociobiol.* **10**, 39–51 (1982).
14. Clutton-Brock, T. H. *The Evolution of Parental Care* (Princeton Univ. Press, Princeton, 1991).
15. Cowley, D. *et al.* The impact of maternal uterine genotype on postnatal growth and adult body size in mice. *Genetics* **122**, 193–203 (1989).
16. Wolf, J. B. & Brodie, E. D. The coadaptation of parental and offspring characters. *Evolution* **52**, 299–308 (1998).
17. Wolf, J. B. Gene interactions from maternal effects. *Evolution* **54**, 1882–1898 (2000).
18. König, B. Cooperative care of young in mammals. *Naturwissenschaften* **84**, 95–104 (1997).

Acknowledgements We thank R. Bshary, T. Coulson, B. Keverne, R. Kilner and S. Rands for comments and discussion. R.H. is supported by the Department of Zoology and Trinity College, Cambridge and The German Academic Exchange Service.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.H. (e-mail: rh244@cam.ac.uk).

A spatially organized representation of colour in macaque cortical area V2

Youpeng Xiao^{*†}, Yi Wang^{*} & Daniel J. Felleman^{*}

^{*} Department of Neurobiology and Anatomy, University of Texas–Houston Medical School, Texas 77030, USA

Neurons responding selectively to different colours have been found in various cortical areas in macaque monkeys¹; however, little is known about whether and how the representation of colour is spatially organized in any cortical area. Cortical area V2 contains modules that respond preferentially to chromatic modulation, which are located in thin cytochrome oxidase stripes^{2–4}. Here we show that within and beyond these modules, gratings of different colours produce activations that peak at different locations. Optical recording of intrinsic signals revealed that the peak regions of the responses to different colours were spatially organized in the same order as colour stimuli are arranged in the DIN (German standard colour chart) colour system. Nearby regions represented colours of a similar hue. We found that the set of colour-specific regions formed 0.07–0.32-mm-wide and approximately 1.3-mm long bands that varied in shape from linear to nearly circular. Our finding suggests that thin stripes in V2 contain functional maps where the colour of a stimulus is represented by the location of its response activation peak.

We identified ‘colour-preferring’ modules in cortical area V2 in differential images resulting from the subtraction of the responses to black/white gratings from those elicited by isoluminant red/green gratings. Figure 1b illustrates such a differential image, in which the dark region marked by a white cross represents a colour-preferring module. The identified colour-preferring modules did not overlap with the orientation-selective modules revealed in separate sets of differential images (see Supplementary Fig. 1), which is consistent with previous results^{2–4}.

After identifying a colour-preferring module at a low spatial resolution (22 $\mu\text{m pixel}^{-1}$), we imaged the activations evoked by different colours in a region centred on this module at a high resolution (8.1 $\mu\text{m pixel}^{-1}$). Most of the visual stimuli that we used were static, isoluminant colour/grayscale gratings with low spatial frequency presented at two spatial phases and two pairs of orthogonal orientations. Figure 1a illustrates the CIE (Commission Internationale de l’Eclairage)–xy chromaticity coordinates of the colours that we used. As defined in the DIN colour system, the hue of a colour is determined by the direction of the line connecting the colour and the white point (CIE source C) in the CIE diagram, whereas the length of the line determines the saturation.

Figure 1c–e shows the single-condition images evoked by three different colours in the same cortical region shown in Fig. 1b. In each image, the outlined patches correspond to regions significantly activated by the presented colour ($P < 0.005$), except for those along blood vessels. Complete sequences of the images corresponding to different time points before and during the presentation of different colour stimuli are shown in Supplementary Fig. 2. A comparison of the activation patterns across these images indicates that different colours activated partially overlapping cortical regions and that these activations peaked at different locations. To compare these activations within the significantly activated regions, we first outlined the peak regions of the activation at 75% of the maximal response to each colour. The maximal response— $(-\Delta R/R) \times 10^4$, where R indicates reflectance—to different colours across all cases

[†] Present address: Department of Ophthalmology, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, New York 10029, USA.

was $3.16 (\pm 0.12 \text{ s.e.m.}, n = 78)$, and this was significantly greater than the maximal noise in the blank images taken during no stimulation ($0.66 \pm 0.12, n = 5, P < 6.2 \times 10^{-7}$; Student's *t*-test). (Examples of the blank images recorded from another hemisphere can be found in Supplementary Fig. 3.) The optical signal in our experiments increased monotonically within the first 3 s of the presentation of most stimuli (Fig. 1i), consistent with the previously reported time course of the optical signal of cortical activation⁵. In Fig. 1i, each curve was derived from the pixel that had the maximal signal across the averaged image elicited by a specific colour.

Next, the outlines corresponding to all of the tested colours were superimposed onto the anatomical image of the cortical surface. Whereas each contour illustrates the cortical region giving the maximal response to each tested colour, different colours produce different response magnitudes. Thus, a different colour stimulus might produce a larger activation within another colour contour. Figure 1f illustrates the outlines derived from two experimental trials separated by 5 h. The vascular pattern was used as a guide to combine or compare results from different trials. Figure 1f demonstrates that the peak regions of the responses to different colours formed a 1.4-mm-long, 0.07–0.21-mm-wide band organized in an L-shape. This band overlapped with and extended beyond the red/green-black/white colour-prefering module, which was 0.52 mm long and 0.28 mm wide. The co-localization between the band of peak responses and the red/green-black/white colour-prefering module was observed in all five hemispheres studied. To illustrate clearly the relationships between the different peak activation

regions, the framed region in Fig. 1f was enlarged and is shown in Fig. 1g. This panel shows that the regions of peak responses to different colours were spatially organized in the same arrangement as that of the colours in the DIN colour system. These peak regions formed a band along which nearby regions represented colours of similar hues. The whole band includes peak responses to all tested colours that span the full range of hue.

We imaged seven cortical regions spanning $2.79 \times 2.07 \text{ mm}$ centred on a red/green-black/white colour-prefering module in V2 of five hemispheres of four animals. In five of the imaged regions, the locations of the peak responses to different colours were organized in the same order as described above (Figs 1g, 2a, c, e and 3b). In most experiments, only a subset of the colours marked in Fig. 1a was tested, and the resulting map of peak responses includes one or two outlines for each tested colour. The regions of the peak responses formed 70–320- μm -wide bands with various shapes, ranging from linear (Fig. 2a) to nearly circular (Fig. 2e). Along these bands, nearby regions represent colours of similar hues, and a full cycle of colours is represented by an approximately 1.3-mm-long segment along the bands. In another imaged region that included two colour-prefering modules (Fig. 2h), the peak regions were also organized according to the hues of the colour stimuli. However, in both the anterior end and the middle of the band, a peak response to a green colour stimulus was missing. Only in one imaged region (not shown) was there a large overlap between regions of peak responses to all different colours, and these regions displayed no clear pattern of spatial organization. Overall, in five

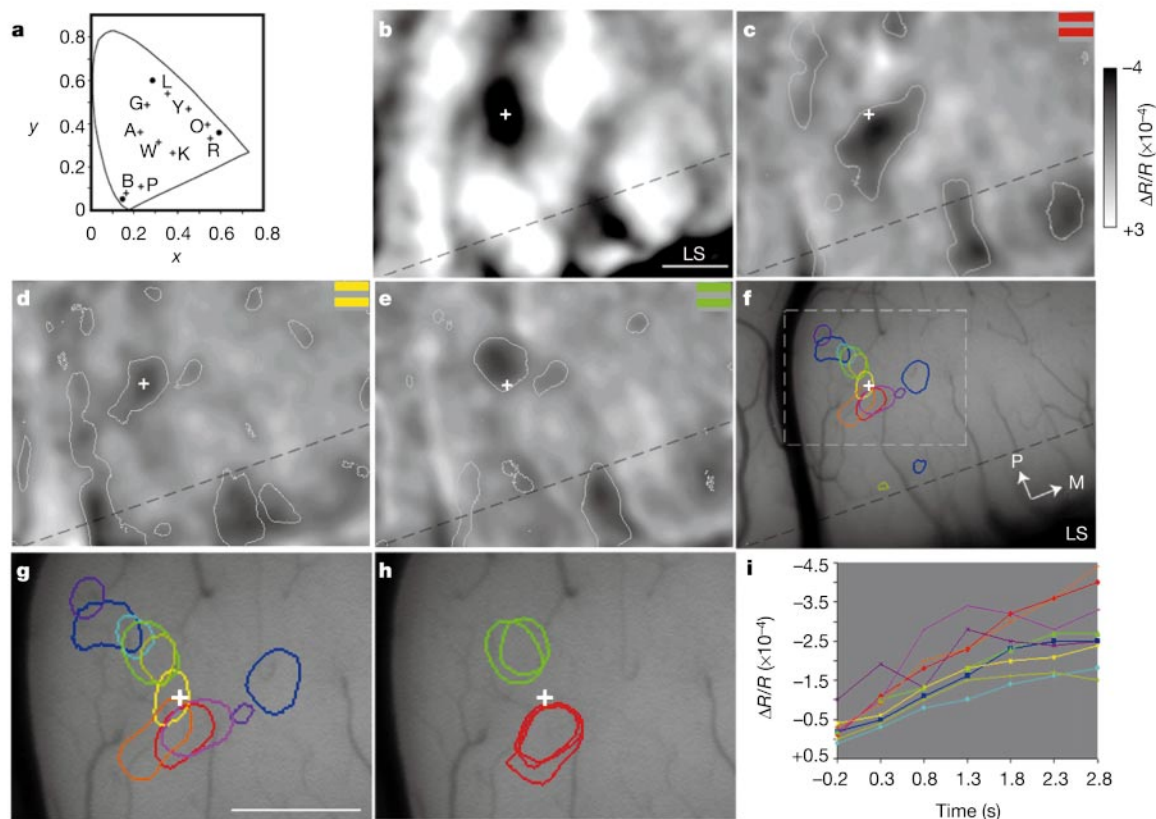


Figure 1 A colour-specific band in V2. **a**, CIE (Commission Internationale de l'Eclairage)-xy chromaticity of the stimulus colours (crosses) and the cathode-ray tube phosphors (circles). A, aqua; B, blue; G, green; K, pink; L, lime; O, orange; P, purple; R, red; W, white and grey; Y, yellow. **b**, Differential image showing a colour-prefering module with its centre marked by a cross. The dashed line marks the excluded region near the lunate sulcus (LS) containing extensive vascular artefacts. **c–e**, Single-condition images evoked

by red (**c**), yellow (**d**) and green (**e**) gratings, with the regions of significant activation outlined. **f**, Outlines of peak responses to different colours. **g**, Enlarged view of the framed region in **f**. **h**, Red and green outlines from separate trials. **i**, Optical signals at different time points before and during the presentations of different colours. M, medial; P, posterior; R, reflectance. Scale bar, 0.5 mm. Case m0103-right (animal number and hemisphere).

out of seven imaged regions, the peak responses to different colours were spatially organized in the order of the hues of the colour stimuli. By chance, the probability of this organization is 1.7%. This probability is estimated by randomly ranking six colours—red, yellow, green, aqua, blue and purple—that were tested in all cases, and comparing each ranking with the order of the corresponding hues in the DIN colour system. When these six colours were ranked according to the locations of their maximal activations along the band of peak responses, the ranking completely followed the order of the hue in 71% of the bands, which is significantly higher than the percentage predicted by a random ranking ($P < 10^{-15}$, χ^2 test). Thus, the observed organization of the peak responses in the order of the hue had a biological origin.

To test the reproducibility of the locations of the peak responses to the same colours, a subset of colours was tested more than once in different trials separated by 2–5 h. The three red outlines in Fig. 1h were derived from two trials conducted 5 h apart, including one trial in which two independent series of images were taken under the red stimulation. The two green outlines in Fig. 1h were from two trials separated by 3 h. On average, the locations of the peak response across repetitions of the stimulus varied by $47.7 \mu\text{m}$ ($10.8 \mu\text{m}$ s.e.m., $n = 8$).

Similar reproducibility was observed in other cases in which different aspects of the stimulus were varied (Fig. 2a–g). In the trial shown in Fig. 2b, four colours were tested that were a subset of the stimuli tested in the trial shown in Fig. 2a. In both trials, the colour gratings were not isoluminant. The trial shown in Fig. 2d tested gratings of a spatial frequency ($2 \text{ cycles degree}^{-1}$) higher than that tested in another trial (Fig. 2c; $0.25 \text{ cycles degree}^{-1}$). Except for blue, each colour evoked peak responses at nearly the same locations in both trials, suggesting that the co-localization between the peak responses and the colour-prefering module does not depend on the spatial frequency of the stimulus. This independence is confirmed by comparing the results of a trial using uniform colours (Fig. 2g) to a trial using isoluminant gratings (Fig. 2e).

To determine whether the shift in the location of the peak activation for each colour was statistically significant, we compared the absolute displacements of these peaks in two cortical hemispheres in which we had exact repetitions in two separate trials of at least four colour stimuli (Fig. 2a, h). Overall, the cortical locations

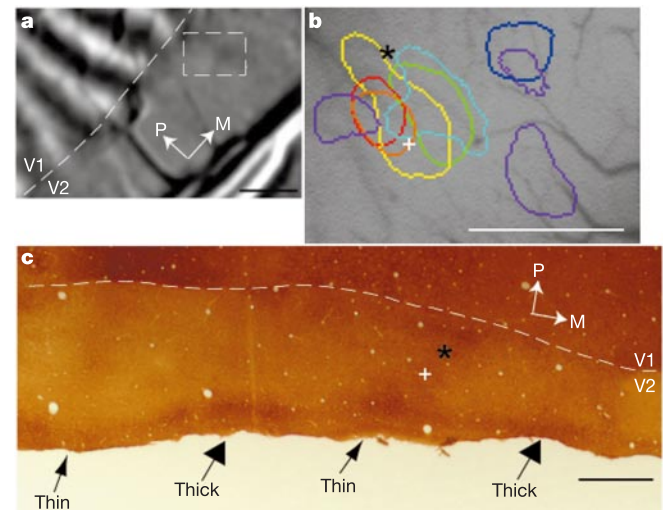


Figure 3 Localization of a colour-specific band. **a**, The V1–V2 border determined by the ocular dominance columns in V1. **b**, A band of peak responses within the framed V2 region in **a**. **c**, A section stained for cytochrome oxidase showing the co-localization of the colour-specific band with a thin stripe. A plus sign indicates the centre of a colour-prefering module; an asterisk indicates the injection site of a tracer. Scale bar, 1 mm, **a**, **c**; 0.5 mm, **b**. Case w04r05-right.

of these response peaks were significantly displaced from each other ($P < 0.005$, Student's t -test), except for the displacement between yellow and green in one case (Fig. 2a).

In all cortical hemispheres studied, the band of the peak responses to different colours was located in area V2. In the hemisphere shown in Fig. 1, this conclusion is based on the patterns of the colour-prefering modules and the orientation-selective columns (Supplementary Fig. 1), the distinctive pattern of blood vessels in V1, and the size of the receptive fields along the band. In all other hemispheres, the ocular dominance columns visualized in V1 were used to identify the V1–V2 border. Figure 3a illustrates the ocular dominance columns in V1 as revealed by the differential image of right eye versus left eye stimulation. The cortical region

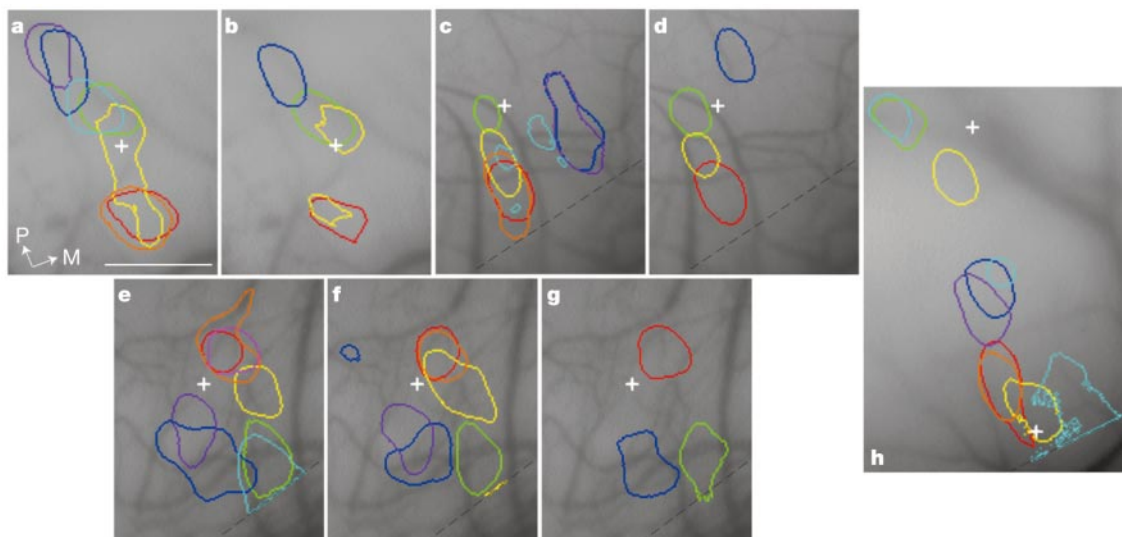


Figure 2 Bands of peak responses to different stimuli. **a**, **b**, Outlined peak responses to luminance-varying gratings in case m0005-left. **c**, **d**, Peak responses to isoluminant gratings of different spatial frequencies (**c**, $0.25 \text{ cycles degree}^{-1}$; **d**, $2 \text{ cycles degree}^{-1}$). Case w02110-left. **e**–**g**, Peak responses to isoluminant (**e**) and luminance-varying (**f**) gratings and uniform colours (**g**). Case m0005-right. **h**, Peak responses to luminance-varying gratings from a region 4 mm lateral to that in **a**. A plus sign indicates the centre of a colour-prefering module. Scale bar, 0.5 mm.

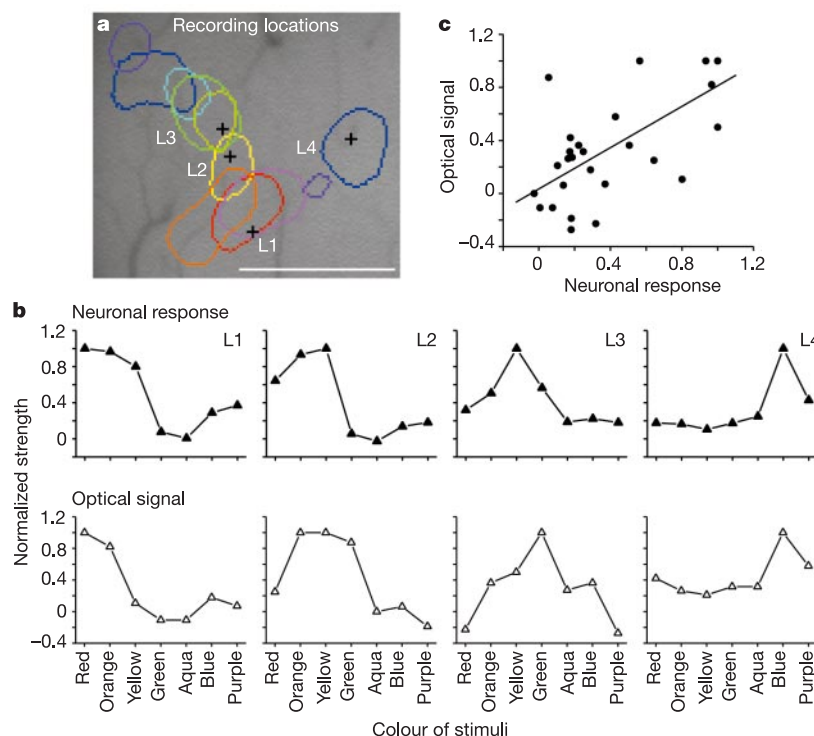


Figure 4 Comparison between neuronal and optical signals. **a**, Four electrode recording locations (L1–4, marked by crosses) in the colour-specific band shown in Fig. 1g. **b**, Colour-tuning profiles of normalized neuronal (top row) and optical (bottom row) signals

at each location. **c**, The scatter plot for each pair of optical (ordinate) and neuronal (abscissa) signals derived from **b**, showing that the two types of signal are significantly correlated ($r = 0.67$, $P < 0.00005$). Scale bar, 0.5 mm. Case m0103-right.

containing a band of peak responses (Fig. 3b) corresponds to the framed region in Fig. 3a, which was located within V2.

To address the relationship between the location of the band of the peak responses and the V2 cytochrome oxidase compartments directly, an injection of the fluorescent tracer FluroRuby was made next to the band in one hemisphere (asterisk in Fig. 3b). The injection site was then identified on post-mortem sections stained for cytochrome oxidase (Fig. 3c). Comparing Fig. 3b with Fig. 3c indicates that the band of the peak responses was largely confined to a thin stripe in this case. Across all hemispheres studied, the bands of the peak responses were localized together with colour-prefering modules that were previously found to be localized in the thin cytochrome oxidase stripes^{2–4}. The extension of these bands along the direction orthogonal to the thin stripes was less than 0.82 mm, comparable to the width of the thin stripes (about 0.74 mm) in this and a previous study⁶. These observations indicate that the bands formed by the peak responses to different colours are confined largely to the thin stripes. However, the band of peak activations to colour stimuli does not occupy the full width or length of a cytochrome oxidase stripe. Within the imaged regions of V2, we estimate that a colour-specific band occupies less than 60% of the area of a thin stripe.

To verify that the optical signal reported here has a neuronal basis, single- or multi-unit activity was recorded from supragranular cortical layers along four vertical penetrations in the case shown in Fig. 1 (Fig. 4a). For each unit, the response to a colour was normalized by the maximal response across all tested colours. The normalized responses of all units along each penetration were averaged and further normalized. These responses to different colours were plotted onto a neuronal colour-tuning profile for each penetration (Fig. 4b, top row). To compare directly the optical and neuronal signal, a similar normalization was applied to the optical signal. At each location of electrode penetration, the optical signal in each single-condition image was normalized by the

maximal signal across different images. For each location, the normalized optical signals corresponding to different colours were plotted onto an optical colour-tuning profile (Fig. 4b, bottom row). Figure 4b demonstrates that at each of the four recording locations, the neuronal and optical signals had very similar colour selectivity. Across all four locations, the two types of signal were significantly correlated ($r = 0.67$, $P < 0.00005$; Fig. 4c), suggesting that the optical signal recorded in our study reflected the local neuronal activity.

Numerous studies have found colour-selective cells in cortical area V2 (refs 4, 7–15). Specifically, the presence in V2 of hue-selective neurons that have narrow selectivity to hue suggests an important role of V2 in the representation of hue¹⁴. In most of these studies, the colour-selective neurons were most common in the thin stripes, consistent with the finding in another study that spatially diffuse variations in colour produce high uptake of ¹⁴C-2-deoxy-D-glucose confined largely in thin stripes⁶. However, it was unknown at that time whether these colour-selective neurons were organized to form a systematic representation of colour. Ts'o and co-workers reported that in the thin stripe, neurons encountered in the same vertical electrode penetration preferred the same wavelength, and the preferred wavelength differed at different cortical locations². Consistent with these studies, we report here that the cortical responses in V2 to different colours peak at different locations in the thin stripes, and that the peak responses are spatially arranged in the order of the hues of the colour stimuli. Our results suggest that in the V2 thin stripes, the colour-selective neurons are organized into maps in which the colour of a stimulus is represented by the location of the peak response to the stimulus. □

Methods

Physiological preparation

Experiments were conducted on four juvenile female macaque monkeys (*Macaca fascicularis*). All procedures were approved by the local institutional Animal Welfare Committee and were in full compliance with the Society for Neuroscience guidelines for

the use of laboratory animals. Briefly, monkeys were anaesthetized with sufentanil citrate ($6\text{--}12\text{ }\mu\text{g kg}^{-1}\text{ h}^{-1}$; intravenous (i.v.) injection) and paralysed with Pavulon (pancuronium bromide; $0.05\text{ mg kg}^{-1}\text{ h}^{-1}$; i.v.). CO_2 and body temperature were maintained at 4.0% and 37.5°C , respectively. Each eye was brought into focus on the screen of a Trinitron monitor using custom-fit contact lenses. On two animals with apparent binocular divergence, the two eyes were brought into convergence by a prism. The whole screen occupied a visual field of $19^\circ \times 14^\circ$ that covered the visual field of the recorded portions of V2 ($2\text{--}7^\circ$ along the vertical meridian). In one animal, anatomical tracers were injected into V2. Its post-mortem occipital operculum was sectioned in the tangential plane and stained for cytochrome oxidase and tracers. A detailed description of these methods has been published previously⁷.

Optical recording

The intrinsic optical signal was recorded using a slow-scan CCD (charge-coupled device) array camera (Photometrics, CH 250 0206) that was focused $0\text{--}300\text{ }\mu\text{m}$ below the surface by a tandem lens system. The cortical surface was illuminated by light at $630 \pm 15\text{ nm}$ wavelength. In each series of frames, four 103-ms -long frames taken $1.3\text{--}2.8\text{ s}$ after the onset of each stimulus were used as response frames. Two background frames were taken at 0.7 and 0.2 s before stimulus onset. Each stimulus, viewed binocularly by the animal, lasted for 3 s and was separated by 10 s of uniform grey on the screen. In each trial, 50 repetitions of each stimulation condition were interlaced pseudo-randomly with other conditions, and the corresponding frames were averaged. A differential image between two stimulation conditions was calculated by subtracting the response frame of one condition from that of the other condition, divided by the first background frame. A single-condition image of one stimulation condition was calculated by dividing the corresponding response frame with the first background frame. For each stimulation condition, a control image was calculated by dividing the second background frame with the first background one. In each trial, four control images derived from four different stimulation conditions were used for the statistical analysis. In each functional image, a pair of gaussian spatial filters (s.d. = 49 and $326\text{ }\mu\text{m}$, respectively) was used to remove the high and low spatial frequency noises. The four single-condition images corresponding to four different time points under the same stimulation condition were then averaged to obtain the final single-condition image for each condition. For each pixel, the intensity values in these four single-condition images were compared with those in the four control images by the Student's *t*-test, to determine the significance level of the activation. In two animals, a series of frames was taken when the screen was kept at a constant uniform grey (blank presentation). The blank presentations were pseudo-randomly interlaced with the stimulus presentations. When the single-condition images during the blank presentations were used as controls, the regions significantly activated by each stimulus were largely the same as those derived from the above analysis, which used the second background images as controls (Supplementary Fig. 3).

To identify the colour-preferring modules, we used isoluminant red/green gratings (sinusoidal, $0.25\text{ cycles degree}^{-1}$, drifted at 1 cycle s^{-1}), high-contrast achromatic gratings (sinusoidal, $2.0\text{ cycles degree}^{-1}$, 2.0 cycles s^{-1} , 100% luminance contrast) and low-contrast achromatic gratings (sinusoidal, $0.25\text{ cycles degree}^{-1}$, 2.5 cycles s^{-1} , 7% luminance contrast). They all had an average luminance of 10 cd m^{-2} and four orientations (0° , 90° , 45° and 135°). The two colours in the red/green gratings corresponded to the colours of the red and green phosphors of the cathode-ray tube, respectively (see Fig. 1a). The colour-preferring modules that we identified appeared dark in the differential image that subtracted the responses to high-contrast achromatic gratings from those to isoluminant gratings, but not in the differential image that subtracted the responses to high-contrast gratings from those to low-contrast achromatic gratings.

To study the cortical activation evoked by a single colour, we used static, colour/grey gratings (square wave, $0.25\text{ cycles degree}^{-1}$, two alternating spatial phases, 0° , 90° , 45° , 135°), or uniform colours covering the full screen (10 cd m^{-2}). CIE-xy chromaticity coordinates of the tested colours are: red (0.55 , 0.33), orange (0.54 , 0.40), yellow (0.45 , 0.47), lime (0.35 , 0.54), green (0.27 , 0.49), aqua (0.23 , 0.36), blue (0.16 , 0.08), purple (0.23 , 0.11), pink (0.38 , 0.27), and white and grey (0.31 , 0.31). The uniform grey between stimuli and the grey in each grating had a luminance of 10 cd m^{-2} . In isoluminant gratings, all colours had a luminance of 10 cd m^{-2} . In luminance-varying gratings, the luminance of different colours were (in cd m^{-2}): red, 14 ; orange, 18 ; yellow, 28 ; green, 19 ; aqua, 22 ; blue, 8 ; purple, 11 . These values are similar to those used by ref. 9 so that they reflect the difference in luminance of the colours described by different names of a natural language.

Electrode recording

Electrode penetrations were made vertically with the guide of functional maps. All units were recorded between 0 and 1 mm below cortical surfaces. After the receptive field of a single- or multi-unit was plotted, flashed colour bars or drifting colour/grey gratings ($1.5\text{ cycles degree}^{-1}$, 1 cycle s^{-1}) were presented in the receptive field for 1.28 s every 3.28 s . Net neuronal responses were obtained by subtracting the spontaneous firing rate during 1 s before the presentation of each stimulus from that during the presentation. Ten repetitions of each stimulus were interlaced pseudo-randomly with other stimuli, and their net neuronal responses were averaged to obtain the mean response to the stimulus.

Received 30 September; accepted 26 November 2002; doi:10.1038/nature01372.

- Komatsu, H. Mechanisms of central colour vision. *Curr. Opin. Neurobiol.* **8**, 503–508 (1998).
- Ts'o, D. Y., Roe, A. W. & Gilbert, C. D. A hierarchy of the functional organization for colour, form and disparity in primate visual area V2. *Vision Res.* **41**, 1333–1349 (2001).
- Xiao, Y., Zych, A. & Felleman, D. J. Segregation and convergence of functionally defined V2 thin stripe and interstripe compartment projections to area V4 of macaques. *Cereb. Cortex* **9**, 792–804 (1999).
- Roe, A. W. & Ts'o, D. Y. Visual topography in primate V2: Multiple representation across functional

stripes. *J. Neurosci.* **15**, 3689–3715 (1995).

- Malonek, D. & Grinvald, A. Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: Implications for functional brain mapping. *Science* **272**, 551–554 (1996).
- Tootell, R. B. & Hamilton, S. L. Functional anatomy of the second visual area (V2) in the macaque. *J. Neurosci.* **9**, 2620–2644 (1989).
- Burkhalter, A. & Van Essen, D. C. Processing of colour, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *J. Neurosci.* **6**, 2327–2351 (1986).
- Levitt, J. B., Kiper, D. C. & Movshon, J. A. Receptive fields and functional architecture of macaque V2. *J. Neurophysiol.* **71**, 2517–2542 (1994).
- Yoshioka, T., Dow, B. M. & Vautin, R. G. Neuronal mechanisms of colour categorization in areas V1, V2 and V4 of macaque monkey visual cortex. *Behav. Brain Res.* **76**, 51–70 (1996).
- Hubel, D. H. & Livingstone, M. S. Segregation of form, colour, and stereopsis in primate area 18. *J. Neurosci.* **7**, 3378–3415 (1987).
- DeYoe, E. A. & Van Essen, D. C. Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature* **317**, 58–61 (1985).
- Peterhans, E. & von der Heydt, R. Functional organization of area V2 in the alert macaque. *Eur. J. Neurosci.* **5**, 509–524 (1993).
- Gegenfurtner, K. R., Kiper, D. C. & Fenstemaker, S. B. Processing of colour, form, and motion in macaque area V2. *Vis. Neurosci.* **13**, 161–172 (1996).
- Kiper, D. C., Fenstemaker, S. B. & Gegenfurtner, K. R. Chromatic properties of neurons in macaque area V2. *Vis. Neurosci.* **14**, 1061–1072 (1997).
- Moutoussis, K. & Zeki, S. Responses of spectrally selective cells in macaque area V2 to wavelengths and colours. *J. Neurophysiol.* **87**, 2104–2112 (2002).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank L. Cleary, E. Kaplan, J. Maunsell and J. Krauskopf for their suggestions on the manuscript. We also thank A. Zych and X. Huang for computer programming, Q. Huang for technical support and J. Chen for discussion. This research was supported by an individual grant from the National Eye Institute to D.J.F. and a core grant from the National Eye Institute to the University of Texas, Health Science Center at Houston.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.X. (e-mail: youping@camelot.mssm.edu).

Inhibitory feedback required for network oscillatory responses to communication but not prey stimuli

Brent Doiron[†], Maurice J. Chacron[†], Leonard Maler[†], André Longtin[‡] & Joseph Bastian[‡]

^{*} *Physics Department, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada*
[†] *Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Ontario K1H 8M5, Canada*
[‡] *Department of Zoology, University of Oklahoma, Norman, Oklahoma 73019, USA*

Stimulus-induced oscillations occur in visual^{1,2}, olfactory^{3–6} and somatosensory⁷ systems. Several experimental^{2,3,5} and theoretical^{8–13} studies have shown how such oscillations can be generated by inhibitory connections between neurons. But the effects of realistic spatiotemporal sensory input on oscillatory network dynamics and the overall functional roles of such oscillations in sensory processing are poorly understood. Weakly electric fish must detect electric field modulations produced by both prey (spatially localized)¹⁴ and communication (spatially diffuse)¹⁵ signals. Here we show, through *in vivo* recordings, that sensory pyramidal neurons in these animals produce an oscillatory response to communication-like stimuli, but not to prey-like stimuli. On the basis of well-characterized circuitry¹⁶, we construct a network model of pyramidal neurons that predicts that diffuse delayed inhibitory feedback is required to achieve oscil-

latory behaviour only in response to communication-like stimuli. This prediction is experimentally verified by reversible blockade of feedback inhibition that removes oscillatory behaviour in the presence of communication-like stimuli. Our results show that a sensory system can use inhibitory feedback as a mechanism to 'toggle' between oscillatory and non-oscillatory firing states, each associated with a naturalistic stimulus.

Organisms must discriminate between sensory inputs with varying spatiotemporal structure. Electric fish offer a clear example because prey and communication signals differ greatly in their spatial extent. Sensory systems have evolved architectures to process such signals efficiently. Feedback connections are common in neural systems, especially at the thalamo-cortical level, where they control sensory transmission². Feedback is also a prominent feature of brainstem electrosensory circuitry, and its organization parallels that of cortico-thalamic pathways¹⁶. We show here how inhibitory feedback allows electrosensory lateral line lobe (ELL) pyramidal neurons¹⁶ to switch between non-oscillatory and oscillatory states, each associated with prey-like or communication-like stimuli, respectively.

The electric fish *Apteronotus leptorhynchus* produces an electric field through a rhythmic electric organ discharge (EOD). Electroreceptors respond to amplitude modulations of this field produced by conspecifics or objects. These afferents encode amplitude modulations¹⁷ and synapse onto ELL pyramidal neurons^{16,18}. Two stimulus geometries, 'local' and 'global', were used to activate receptor afferents and pyramidal neurons (see Methods). Local geometry applies the stimulus to a small fraction of the fish's skin, stimulating only a portion of the centre of the receptive field of a pyramidal neuron from which we record (Fig. 1a, left). This produces localized amplitude modulations of the fish's own EOD that are spatially similar to those produced by small aquatic invertebrate prey¹⁴. Global geometry, however, applies a bilateral stimulus homogeneously to the whole fish, stimulating both the centre and surround receptive fields of pyramidal neurons (Fig. 1b, left). This produces amplitude modulations that are spatially similar to communication signals¹⁵. Large objects (obstacles, root masses) will produce spatially extensive amplitude modulations, but these are expected to be heterogeneous. These fish routinely give communication responses¹⁵ in response to global stimuli with gaussian temporal structure (data not shown), showing that these effectively mimic communication stimuli.

Figure 1 shows a typical pyramidal cell response recorded *in vivo* when band-limited gaussian EOD amplitude modulations (0–40 Hz) were applied locally or globally ($n = 15$). With local stimulation, the spike time autocorrelation¹⁹, $A(\tau)$, was positive at small values of the lag, τ (after a negative correlation owing to refractoriness), and histograms of the interspike interval (ISI) showed a single peak (Fig. 1a, middle and right). $A(\tau)$ shows that the spikes are not independent of each other; rather, there is a tendency to produce high-frequency (~ 100 -Hz) clusters of spikes²⁰. By contrast, when the stimulus was applied globally, a damped oscillation in $A(\tau)$ and two distinct peaks in the ISI histogram appeared, suggesting oscillatory (and/or bursting) behaviour (Fig. 1b, middle and right); no such structure is observed for local stimulation. It is this transition from non-oscillatory to oscillatory behaviour contingent on the switch from local to global stimulation that is the focus of our study.

To verify the presence of oscillatory behaviour and to compare quantitatively the discharge patterns under both stimulus geometries, we computed the power spectral densities (PSDs) of the spike train during both global and local stimulation (Fig. 1d). There was a significant peak in the PSDs during global stimulation (32.41 ± 3.92 Hz; $n = 15$) that was absent during local stimulation, demonstrating that the spike train is oscillatory only during global stimulation. Using the PSDs, we defined an 'oscillation index' (see Methods); this index was larger under global ($10.12 \pm$

$3.68 \text{ spikes}^2 \text{ s}^{-1}$) than under local ($2.28 \pm 0.62 \text{ spikes}^2 \text{ s}^{-1}$) geometries (pair-wise t -test, $P = 0.0034$; $n = 15$). In addition, joint interval histograms computed from responses to global and local stimuli differed markedly. Under local stimulation, the joint interval histograms showed no clear structure; however, with global stimuli, short ISIs (< 15 ms) were followed preferentially by long ISIs (> 15 ms) and vice versa (Fig. 1c). Oscillatory discharge was not observed with local stimulation, even with a contrast up to six times that of the global stimulus, although the peak firing rates were similar to those during global stimulation. These results show that an oscillatory discharge occurs for global but not local stimulation.

The requirement of a global stimulus for oscillatory discharge suggests that the mechanism involves the integration of spatially diffuse inputs. The negative component of $A(\tau)$ at $\tau \approx 15$ ms suggests that delayed inhibitory processes are also involved. ELL circuitry is well described and indicates that feedforward pathways satisfying these criteria are unlikely to exist¹⁶. But ELL pyramidal neurons project to an isthmus structure (n. praeminentialis dorsalis), which in turn provides several feedback projections to the ELL¹⁶. This includes a spatially diffuse delayed (~ 15 ms) inhibitory pathway emanating from a small group of bipolar cells that terminates exclusively on ELL pyramidal cells^{21–23}. Theoretical analysis of models of biological and chemical systems with delayed negative feedback have shown how delays can often support oscillatory dynamics²⁴; however, little is known about how the spatial extent of a driving stimulus can regulate a feedback-mediated oscillation. Thus, we constructed a neural network model of the ELL to test the hypothesis that delayed inhibitory feedback is sufficient for global stimulus-induced oscillations.

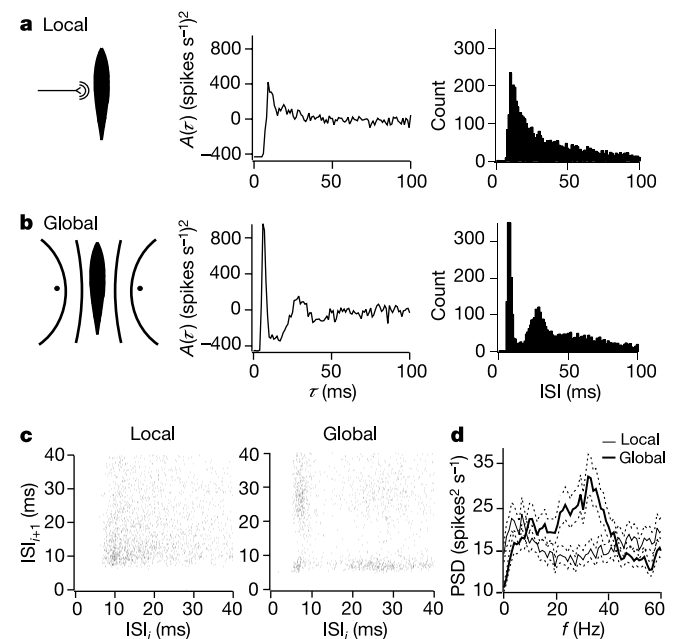


Figure 1 ELL pyramidal neurons show differential responses to local (prey-like) and global (communication-like) stimuli ($n = 15$). **a**, Local stimulation. Left, local stimulus model; a dipole was placed near the skin to stimulate only a part of the pyramidal cell receptive field centre. Middle and right, $A(\tau)$ and ISI histograms of a typical pyramidal cell response to a local stimulus. **b**, Global stimulation. Left, global stimuli model; two Ag^+/AgCl wire electrodes were placed transverse to the fish to produce spatially extensive stimuli. Middle and right, $A(\tau)$ and ISI histograms of pyramidal cell response to a globally applied stimulus. The mean firing rates of the stimulated cell were $21.8 \text{ spikes s}^{-1}$ and $22.8 \text{ spikes s}^{-1}$ for local and global stimulus geometries, respectively. Note that the same cell was used for both stimuli. **c**, Joint interval histograms under local (left) and global (right) stimulation. **d**, The PSDs of the spike train under local and global stimulus geometries. Broken lines surrounding the PSDs are the 95% confidence bands (± 2 s.d.).

In brief, we modelled the pyramidal neurons as a network of leaky integrate-and-fire (LIF) neurons (ref. 19 and see Methods). All neurons transmit their spikes to a central cell population, G . For each spike that it receives, G feeds back an inhibitory synaptic (α -function¹⁹) response to all pyramidal neurons in the network after a fixed time delay, τ_d . The physiological interpretation of G is the integration of pyramidal cell output by bipolar cells of the n. praeminentialis dorsalis and their GABA (γ -aminobutyric acid)-mediated projection back to the ELL. A diagram of the network model connectivity is shown in the left panel of Fig. 2a.

Each LIF neuron was driven by two sources. The first was a noise input, $\eta_i(t)$, with positive bias (the subscript i denotes spatial position), with η being uncorrelated across space. This modelled stochastic spontaneous activity is similar to that seen *in vivo*²⁰. The second input source was band-limited (0–40 Hz) gaussian noise, $S(t)$, mimicking the temporal nature of the electrosensory stimulus used in the experiments. To model local geometry, $S(t)$ was applied to a single neuron, whereas global geometry was modelled by applying $S(t)$ homogeneously to all neurons. With local stimulation, $A(\tau)$ and ISI histograms measured from the neuron receiving $S(t)$ were similar to those observed experimentally, indicating a non-oscillatory state (compare Figs 1a and 2a, middle and right). With global stimulation, however, a transition to an oscillatory response was observed, similar to that observed experimentally (compare Figs 1b and 2b, middle and right). The oscillation index of the model was $8.66 \text{ spikes}^2 \text{ s}^{-1}$ during local stimulation and $15.0 \text{ spikes}^2 \text{ s}^{-1}$ during global stimulation. Joint interval histograms constructed from model spike trains during both local and global conditions were qualitatively similar to those constructed from the data.

These model results were insensitive to parameter heterogeneities in the network resulting from the known variability of pyramidal neurons (ref. 20 and see Methods). In addition, a sixfold increase in the contrast of a local stimulus did not induce oscillatory discharge, similar to experimental findings. We have previously proposed a two-compartmental model of ELL pyramidal cells²⁵ that simulates the single-cell bursting dynamics observed during current injection *in vitro*²⁶. We have also qualitatively reproduced the simulation results presented above (Fig. 2) using a network of these more realistic pyramidal cells (see Supplementary Information).

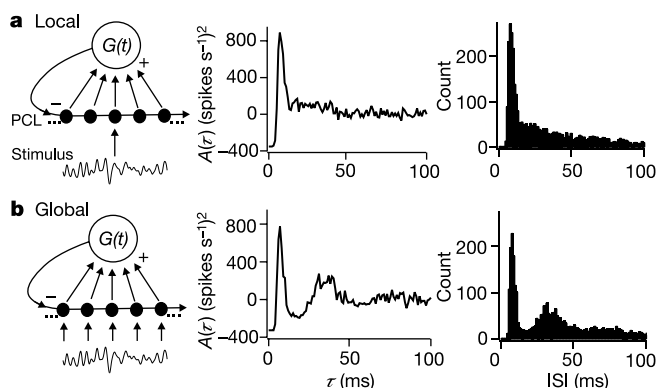


Figure 2 ELL neural network simulations involving global inhibitory feedback show differential responses to local and global stimuli. **a**, Local stimulation. Left, network showing the pyramidal cell layer (PCL) projecting spikes to a kernel, $G(t)$, which feeds back inhibitory responses to the PCL after a delay τ_d . For local stimuli only one cell receives the stimulus (arrow). Middle and right, $A(\tau)$ and ISI histograms of the simulated LIF neuron during local stimulation. **b**, Global stimulation. Left, identical network to that in **a**, but all LIF neurons receive the same stimulus. Middle and right, $A(\tau)$ and ISI histograms of a LIF neuron response to global stimulation. The mean firing rates of the cell were $17.5 \text{ spikes s}^{-1}$ and $16.5 \text{ spikes s}^{-1}$ for local and global stimulus geometries, respectively.

During local stimulation, the only shared input among neurons is from the global inhibitory feedback, G . In spite of this common input, the intrinsic noise sources, $\eta_i(t)$, were of sufficient strength to suppress significant correlations between neurons in the network. This is demonstrated in the model network raster plots (Fig. 3a, top), which show a lack of coincident firing between individual neurons during local input, and by network stimulus histograms (summed activity of all neurons) that weakly fluctuate about a constant value (Fig. 3a, middle). The feedback response, G (Fig. 3a, bottom), is determined by the network firing pattern, as illustrated by the network spike histograms, and is thus also roughly constant.

During global conditions, however, $S(t)$ was distributed homogeneously over the whole network, evoking stimulus-induced correlations between neurons. This allowed for increased network synchronization, as shown in network raster plots (Fig. 3b, top) and network stimulus histograms (Fig. 3b, middle). As a result of this increased correlation, G received near-coincident spikes from across the network, yielding a significant summation of α -function responses. This enhanced response is fed back to all neurons in the network after a delay time, τ_d , causing ‘waves’ of inhibition that suppress network firing (Fig. 3b, bottom). As an inhibitory wave abates, subsequent stimulus-induced coincident firing across the network develops. Thus, the interplay between stimulus-induced correlations and an associated delayed increase of feedback inhibition creates an oscillation in the network activity with period determined by both the axonal delay, τ_d , and the synaptic time constant α .

The simulation results shown in Figs 2 and 3 suggest that the diffuse delayed inhibitory feedback is an integral component of the pyramidal cell oscillatory response to global stimuli. To verify this experimentally, we used local injection of xylocaine, a sodium channel antagonist, to specifically block the inhibitory feedback component of the tractus stratum fibrosum (StF) pathway ($n = 10$; see Fig. 4 and Methods). The effect of this block was to remove temporarily feedback inhibition to the ELL, effectively opening the loop from the n. praeminentialis dorsalis to ELL. Xylocaine injection reversibly blocked oscillatory responses to globally applied stimuli: $A(\tau)$ showed no damped oscillations and the ISI histograms were unimodal (Fig. 4b, c, black traces). However, these neurons

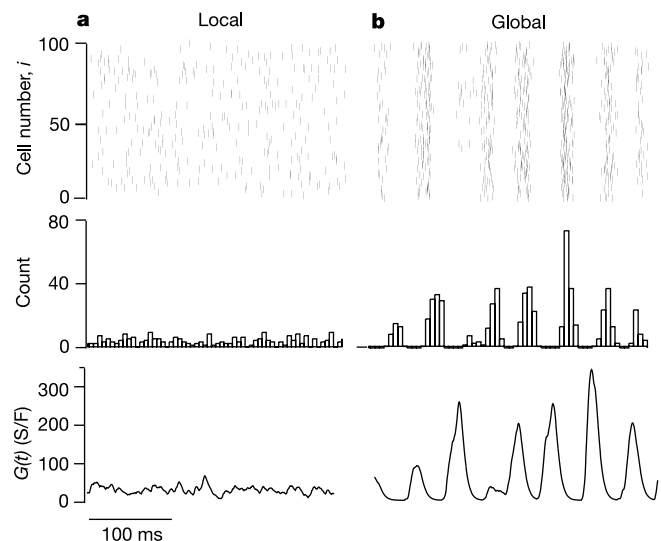


Figure 3 Simulated asynchronous firing and synchronized oscillations during one presentation of a local or global stimulus, respectively. Top, network raster plots under local (**a**) and global (**b**) stimulus with network stimulus histograms shown underneath. Bottom, the total value of the feedback inhibition, $G(t)$, is plotted for the same simulations that gave the spike times.

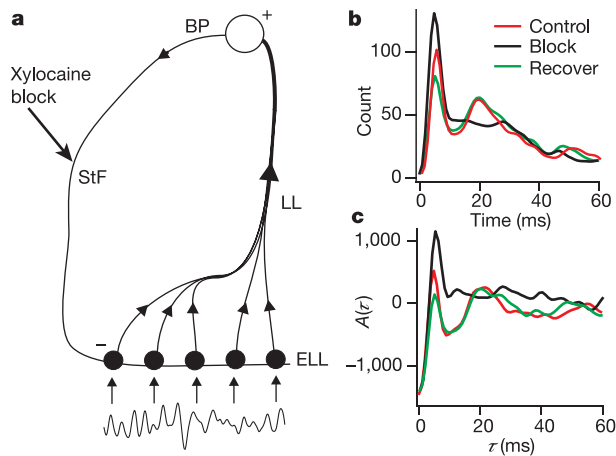


Figure 4 Blockade of the inhibitory component of the StF pathway. **a**, Pyramidal neurons in the ELL project excitatory connections to bipolar cells (BP) of the n. praeminentialis dorsalis through the lateral lemniscus pathway (LL). In turn, the BP cells project back to the ELL through the StF, providing GABA-mediated feedback to the pyramidal neurons. StF axons were blocked by application of xylocaine. **b**, **c**, ISI histograms (**b**) and $A(\tau)$ (**c**) are plotted for control cells (red), during the block (black) and after recovery from xylocaine (green). Histograms and $A(\tau)$ were constructed as in Fig. 1 and smoothed (gaussian-filtered, $\sigma = 5$ ms) to facilitate comparison between states.

showed clear oscillatory behaviour before and after recovery from StF (inhibitory component) blockade (Fig. 4b, c, red traces, control; green traces, recovery). The oscillation indices for control and recovery were similar (pair-wise t -test, $P = 0.353$; $n = 10$) and significantly different from those observed during feedback inhibitory blockade (pair-wise t -test, $P = 0.02$; $n = 10$). These experiments verify the model's prediction that diffuse inhibitory feedback is required for ELL pyramidal neurons to show differential responses to local versus global stimuli.

Our network simulations predict that the synchronization of ELL pyramidal cell firing is an aspect of the oscillatory mechanism involved during global stimulation (Fig. 3). Field potential recordings from the ELL show oscillations during global stimulation (data not shown), consistent with synchronous activity; however, inter-neuron populations in the ELL¹⁶ and the expected stimulus-driven increase of synchronization during global stimulation complicate the interpretation of this observation.

Inhibition is necessary for synchronization in the olfactory system^{3,5} and thalamocortical circuits^{2,8}. Modelling studies have shown that synchronous oscillatory solutions exist when inhibitory feedback gain is strong and inhibitory synaptic timescales are long^{8,9,11,13}, or when there are sufficiently long axonal delays^{9,10,13}. Other studies have shown that network oscillations can occur with constant depolarizing input¹² or with increased mean firing rate of uncorrelated excitatory Poisson inputs¹³. In the latter study, oscillations also occurred with increased spatial spread of uncorrelated inputs distributed to the network. In this study, by contrast, synchronous oscillations were produced not by varying feedback gain, intrinsic cellular properties or mean network depolarization, but by simply changing the spatial extent of a stimulus.

It is known that pyramidal neurons can estimate the time course of prey-like stimuli applied under local, but not global, geometries²⁷. But these cells do produce bursts of spikes in response to specific features of a global stimulus²⁸. These studies suggest that pyramidal cell coding strategies may also depend on the spatial extent of stimuli. We have confirmed that the stimulus encoding abilities of pyramidal cells can be altered during the oscillatory discharge induced by a global stimulus (see Supplementary Information).

We have described a mechanism whereby the spatial extent of stimuli determines the firing behaviour of sensory neurons in the electrosensory system. In addition, these stimuli were spatially similar to those that the fish would encounter in either prey location or communication situations, giving a clear behavioural significance to our results. We have shown that this mechanism requires delayed inhibitory feedback—a very general feature of the nervous system^{2,29}. To our knowledge, this is the first demonstration of how delayed inhibitory feedback enables neurons in a sensory channel to toggle between two distinct firing states, with each state connected to a behaviourally relevant stimulus. □

Methods

Physiology

Extracellular single-unit recordings from ELL pyramidal neurons were identified on the basis of recording depth and response patterns²⁷. Random amplitude modulations (RAMs) of the EOD were produced by adding an amplitude-modulated sinusoidal waveform to the continuing discharge of the fish. The sinusoid was phase locked to the EOD with a period roughly equal to that of the EOD. RAMs were produced by multiplying the sinusoid by zero-mean band-limited gaussian noise (0–40 Hz, eighth order Butterworth filter). For local geometry, we applied stimuli through a small dipole, tip spacing of 2 mm, positioned at a site in the cell's receptive field centre 2–3 mm lateral to the skin. Global geometry stimuli were applied homogeneously through two electrodes 19 cm lateral to either side of the fish (Fig. 1b, left). The typical global stimulus amplitude was $250 \mu\text{V cm}^{-1}$, and the magnitude of the local stimulus was adjusted so that electroreceptor afferent responses were equivalent for either geometry²⁷. Histograms and autocorrelation bin widths were 1 ms and were computed from about 4,000 spikes in all cases. ELL pyramidal cells occur in both basilar and non-basilar varieties¹⁸; our results include both subtypes. We limited recordings to pyramidal neurons from lateral and contralateral ELL segments with mean spontaneous firing frequencies greater than 15 Hz.

Reversible block of the tractus StF, which conveys descending excitatory and inhibitory inputs to pyramidal neurons²¹, was achieved through micropressure ejection of the local anaesthetic xylocaine. A xylocaine filled micropipette (5–10- μm tip diameter) carrying a bipolar pair of 25- μm stainless steel stimulation wires was positioned rostral to the ELL at a depth where electrical stimulation (0.1 ms, 15–25 μA current pulses) evoked the characteristic StF field potential in the ELL²². Because the inhibitory component of the StF lies at its ventral-most aspect just dorsal to the lateral lemniscus (ELL efferents)²¹, the pipette was positioned just dorsal to a site where antidromic activation of ELL pyramidal neurons occurred. Ejection of xylocaine (2% concentration; 100-ms pulses at 50 p.s.i.) reduced the StF field potential, verifying blockade; the loss of pyramidal cell oscillatory responses paralleled this reduction. Blockade of the dorsal StF, which conveys topographical excitatory feedback^{16,22}, eliminated the StF field potential but did not prevent global stimulus-induced pyramidal cell oscillations (data not shown). All procedures were in accordance with the University of Oklahoma animal care and use guidelines. The oscillation index was defined as the difference between the maximum and the minimum value of the PSD between 20 and 40 Hz.

Network simulations

We model the ELL pyramidal cell layer as a network of N LIF model neurons¹⁹. The membrane potential of the i th neuron is labelled $V_i(t)$. When $V_i(t)$ crosses a spike threshold θ , it is reset to the potential (V_{reset}) and a spike is emitted from neuron i . Between spike times ($V_i(t) < \theta$), the dynamics are governed by the following stochastic delay-differential system:

$$\frac{dV_i}{dt} = \frac{-V_i}{\tau_m} + \eta_i(t) + B_i + S_i(t) - G(t - \tau_d)(V_i - V_r) \quad (1)$$

$$G(t) = \frac{g}{N} \sum_{j=1}^N \sum_{m=0}^{M_j(t)} \frac{(t - t_{jm})}{\alpha} \exp\left(1 - \frac{(t - t_{jm})}{\alpha}\right) \quad (2)$$

τ_m is the passive membrane time constant, and a constant depolarizing bias B_i is set so that, for zero input ($\eta_i = g = S = 0$), each neuron fires periodically ($B_i > \theta/\tau_m$). $\eta_i(t)$ is a zero mean Ornstein–Uhlenbeck process³⁰, with time constant τ_η . The Ornstein–Uhlenbeck processes are independent ($\eta_i(t)$ and $\eta_j(t)$ are not correlated if $i \neq j$) and the autocorrelation of $\eta_i(t)$ is given by

$$\langle \eta_i(t) \eta_i(t + \tau) \rangle = \frac{\sigma^2}{2\tau_\eta} e^{-\tau/\tau_\eta} \quad (3)$$

RAMs were introduced through a zero mean band-limited (0–40 Hz) gaussian noise, $S(t)$, with variance W . Local stimulation was simulated by applying $S(t)$ to only one LIF neuron, whereas for global stimulus geometry $S(t)$ was applied to all neurons.

The direct inhibitory feedback pathway¹⁶ is modelled by the kernel $G(t)$ in equation (2). $G(t)$ receives spikes from the j th neuron, t_{jm} , over the history of neuron j ($0 \leq m \leq M_j(t)$), where $M_j(t)$ is the total spike count of neuron j at time t . $G(t)$ transforms a spike input, t_{jm} , into an α -function conductance¹⁹ with time constant α . $G(t)$ then projects the summation of all inputs uniformly to all neurons in the network after a fixed delay, τ_d . The gain of the pathway is set by the constant $g > 0$, and $G(t)$ multiplies a battery term ($V_i - V_r$) with reversal potential V_r . The LIF neurons do not have local connections among themselves, as observed for ELL pyramidal cells¹⁶.

The intrinsic model parameters were set to $\tau_m = 10$ ms, $\sigma = 5.5$ nA, $B_i = 0.84$ nA (we

assumed a capacitance of 1 nF) and $\tau_\eta = 15$ ms. These were chosen so that for $S(t) = 0$, $A(\tau)$ and ISI histograms were similar to those observed under spontaneous conditions *in vivo*²⁰. We set $N = 100$ to make local and global stimulation significantly distinct. The RAM stimulus variance was $W = 0.238$ nA², which resulted in firing statistics comparable to those of stimulated pyramidal neurons (Fig. 2). The feedback parameters were set to $g = 390$ S F⁻¹, $\alpha = 3$ ms, $V_r = V_{\text{reset}} = 0$ mV, to model the feedback as GABA_A shunting inhibition, previously reported for interactions between bipolar and pyramidal cells²³. The loop delay was set to $\tau_d = 12$ ms, allowing a fit to experimental data (Figs 1 and 2), and is also comparable to previously estimated values²². Equation (1) was integrated by a simple Euler–Maruyama³⁰ scheme with $\Delta t = 0.025$ ms. The effects of parameter heterogeneities were tested by choosing B_i and τ_m from gaussian distributions with mean values as given above. The results were qualitatively similar to the homogeneous case for distributions with coefficient of variations of 0.82 and 1.5 for the B_i and τ_m distributions, respectively.

Received 20 August; accepted 26 November 2002; doi:10.1038/nature01360.

- Gray, C. & Singer, W. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl Acad. Sci. USA* **86**, 1698–1702 (1989).
- Sillito, A. M., Jones, H. E., Gerstein, G. L. & West, D. C. Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature* **369**, 479–482 (1994).
- MacLeod, K. & Laurent, G. Distinct mechanisms for synchronization and temporal patterning of odor-encoding neural assemblies. *Science* **274**, 976–979 (1996).
- Kashiwadani, H., Sasaki, Y. F., Uchida, N. & Kensaku, M. Synchronized oscillatory discharge of mitral/tufted cells with different molecular receptive ranges in the rabbit olfactory bulb. *J. Neurophysiol.* **82**, 1786–1792 (1999).
- Stopfer, M., Bhagavan, S., Smith, B. H. & Laurent, G. Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature* **390**, 70–74 (1997).
- Friedrich, R. W. & Laurent, G. Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. *Science* **291**, 889–894 (2001).
- Ahissar, E. & Vaadia, E. Oscillatory activity of single units in a somatosensory cortex of an awake monkey and their possible role in texture analysis. *Proc. Natl Acad. Sci. USA* **87**, 8935–8939 (1992).
- Destexhe, A., Contreras, D. & Steriade, M. Mechanisms underlying the synchronization action of corticothalamic feedback through inhibition of thalamic relay cells. *J. Neurophysiol.* **79**, 999–1016 (1998).
- Bressloff, P. C. & Coombes, S. Dynamics of strongly coupled spiking neurons. *Neural Comput.* **12**, 91–129 (2000).
- Ernst, U., Pawelzik, K. & Geisel, T. Synchronization induced by temporal delays in pulse-coupled oscillators. *Phys. Rev. Lett.* **74**, 1570–1573 (1995).
- Wang, X.-J. & Rinzler, J. Spindle rhythmicity in the reticularis thalami nucleus: synchronization among mutually inhibitory neurons. *Neuroscience* **53**, 899–904 (1993).
- van Vreeswijk, C. & Hansel, D. Patterns of synchrony in neural networks with spike adaptation. *Neural Comput.* **13**, 959–992 (2001).
- Paulis, Q., Baker, N. B. & Olivier, E. Emergent oscillations in a realistic network: the role of inhibition and the effect of the spatiotemporal distribution of input. *J. Comput. Neurosci.* **6**, 27–48 (1999).
- Nelson, M. E. & MacIver, M. A. Prey capture in the weakly electric fish *Apteronotus leptorhynchus*: sensory acquisition strategies and electrosensory consequences. *J. Exp. Biol.* **202**, 1195–1203 (1999).
- Metzner, W. Neural circuitry for communication and jamming avoidance in gymnotiform electric fish. *J. Exp. Biol.* **202**, 1365–1375 (1999).
- Berman, N. J. & Maler, L. Neural architecture of the electrosensory lateral line lobe: adaptations for coincidence detection, a sensory searchlight and frequency-dependent adaptive filtering. *J. Exp. Biol.* **202**, 1243–1253 (1999).
- Wessel, R., Koch, C. & Gabbiani, F. Coding of time-varying electric field amplitude modulations in a wave-type electric fish. *J. Neurophysiol.* **75**, 2280–2293 (1996).
- Maler, L., Sas, E. K. & Rogers, J. The cytology of the posterior lateral line lobe of high frequency weakly electric fish (Gymnotoidei): differentiation and synaptic specificity in a simple cortex. *J. Comp. Neurol.* **195**, 87–139 (1981).
- Dayan, P. & Abbott, L. F. *Theoretical Neuroscience* (MIT Press, Cambridge, Massachusetts, 2001).
- Bastian, J. & Nguyenkim, J. Dendritic modulation of burst-like firing in sensory neurons. *J. Neurophysiol.* **85**, 10–22 (2001).
- Maler, L. & Mugnaini, E. Correlating γ -aminobutyric acidergic circuits and sensory function in the electrosensory lateral line lobe of a gymnotiform fish. *J. Comp. Neurol.* **345**, 224–252 (1994).
- Berman, N. J., Plant, J., Turner, R. & Maler, L. Excitatory amino acid transmission at a feedback pathway in the electrosensory system. *J. Neurophysiol.* **78**, 1869–1881 (1997).
- Berman, N. J. & Maler, L. Interaction of GABA_A-mediated inhibition with voltage-gated currents of pyramidal cells: computational mechanism of a sensory searchlight. *J. Neurophysiol.* **80**, 3197–3213 (1998).
- Glass, L. & Mackey, M. C. *From Clocks to Chaos* (Princeton Univ. Press, Princeton, New Jersey, 1988).
- Doiron, B., Laing, C., Longtin, A. & Maler, L. Ghostbursting: a novel neuronal burst mechanism. *J. Comput. Neurosci.* **12**, 5–25 (2002).
- Lemon, N. & Turner, R. W. Conditional spike backpropagation generates burst discharge in a sensory neuron. *J. Neurophysiol.* **84**, 1519–1530 (2000).
- Bastian, J., Chacron, M. J. & Maler, L. Receptive field organization determines pyramidal cell stimulus encoding capability and spatial stimulus selectivity. *J. Neurosci.* **22**, 4577–4590 (2002).
- Gabbiani, F., Metzner, W., Wessel, R. & Koch, C. From stimulus encoding to feature extraction in weakly electric fish. *Nature* **384**, 564–567 (1996).
- Contreras, D., Destexhe, A., Sejnowski, T. J. & Steriade, M. Control of spatiotemporal coherence of a thalamic oscillation by corticothalamic feedback. *Science* **274**, 771–774 (1996).
- Kloeden, P. E. & Platen, E. *Numerical Solutions to Stochastic Differential Equations* (Springer, Berlin, 1992).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank J. Lewis, C. Laing, R. W. Turner and M. Higley for reading the manuscript. Funding was provided by the National Science and Engineering Research Council (B.D., M.J.C., A.L.), the Canadian Institutes of Health Research (A.L., L.M.) and the National Institutes of Health (J.B.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.D. (e-mail: bdoiron@science.uottawa.ca).

Fidelity in planar cell polarity signalling

Dali Ma*, Chung-hui Yang†, Helen McNeill‡, Michael A. Simon† & Jeffrey D. Axelrod*

* Department of Pathology, Stanford University School of Medicine, Stanford, California 94305-5324, USA

† Department of Biological Sciences, Stanford University, Stanford, California 94305, USA

‡ Cancer Research UK, London Research Institute, 44 Lincoln's Inn Fields, London WC2A 3PX, UK

The polarity of *Drosophila* wing hairs displays remarkable fidelity. Each of the approximately 30,000 wing epithelial cells constructs an actin-rich prehair that protrudes from its distal vertex and points distally. The distal location and orientation of the hairs is virtually error free, thus forming a nearly perfect parallel array. This process is controlled by the planar cell polarity signalling pathway^{1–4}. Here we show that interaction between two tiers of the planar cell polarity signalling mechanism results in the observed high fidelity. The first tier, mediated by the cadherin Fat⁵, dictates global orientation by transducing a directional signal to individual cells. The second tier, orchestrated by the 7-pass transmembrane receptor Frizzled^{6,7}, aligns each cell's polarity with that of its neighbours through the action of an intercellular feedback loop, enabling polarity to propagate from cell to cell⁸. We show that all cells need not respond correctly to the presumably subtle signal transmitted by Fat. Subsequent action of the Frizzled feedback loop is sufficient to align all the cells cooperatively. This economical system is therefore highly robust, and produces virtually error-free arrays.

A group of signalling molecules, including Frizzled (Fz)⁷, Dishevelled (Dsh)⁹, Flamingo (Fmi)^{10,11}, Van Gogh^{12,13} and Prickle (Pk)¹⁴ mediates planar cell polarity (PCP) signalling in various developing *Drosophila* tissues^{1–4}. In wing cells, these proteins participate in an intercellular feedback loop that generates subcellular asymmetry and directs prehair location^{8,15,16}. Frizzled on the distal side of one cell recruits Dsh to the membrane, thus stabilizing Fz localization, and simultaneously recruits Pk to the proximal side of the adjacent cell, where it prevents Dsh localization (Fig. 1a). A competition therefore occurs between Fz on either side of the intercellular boundary, amplifying small differences in Fz levels to all-or-none differences⁸. Coupled with the observation that cells do not accumulate high Fz levels and construct prehairsts on two sides, this mechanism enables polarity to propagate from cell to cell (Fig. 1b). For example, removing Fz from a clone of cells produces a difference in Fz at the distal boundary of the clone that is reversed from the wild type, and this reversal propagates, resulting in the 'domineering non-autonomy' first described by ref. 6 (Fig. 1b). The question remains, however, of what initiates the imbalance of Fz activity along the proximal–distal axis of each cell in a wild-type wing. We have proposed previously that, in the eye, two

non-classical cadherins, Dachous (Ds)¹⁷ and Fat (Ft), and a type II transmembrane protein, Four-jointed (Fj)¹⁸ work together to provide the Fz-dependent signal with directional cues¹⁹. In the imaginal eye disc, Ds and Fj, expressed in opposing gradients^{19,20}, function through Ft to enhance Fz activity in the R3 photoreceptor cell^{21,22}. These proteins also affect PCP in the wing and abdomen^{23–25}. We have investigated whether Fj, Ds and Ft provide global directional information in the wing.

Several observations suggest that the activities of Fj, Ds and Ft require Fz, and that they bias the orientation of Fz signalling in the wing, as they do in the eye¹⁹. Although *ft* mutant clones produce PCP phenotypes, Ft signalling requires Fz activity, as loss of *ft*

function in large or small clones does not alter the characteristic polarity pattern of *fz* mutant wings (not shown). Fat localization (see below) is not affected in *fz*, *pk* and *fmi* mutant clones (not shown). Furthermore, in the absence of Fj, Ds or Ft, Dsh is membrane associated, and Dsh, Pk and Fmi remain asymmetrically localized (Fig. 1c, d, and data not shown), indicating that Fz continues to signal^{15,16}, although not always in the correct orientation. Similarly, double mutant combinations between *ds* and *fz* or *hs-fz* indicate that Fz functions, but abnormally, in *ds* mutants²³. Therefore, in the wing, as in the eye¹⁹, Fj, Ds and Ft orient the direction of the Fz-mediated intercellular feedback loop.

Four-jointed expression in the wing is graded, with the highest level at the distal margin (Fig. 1e; see also ref. 24). We found that

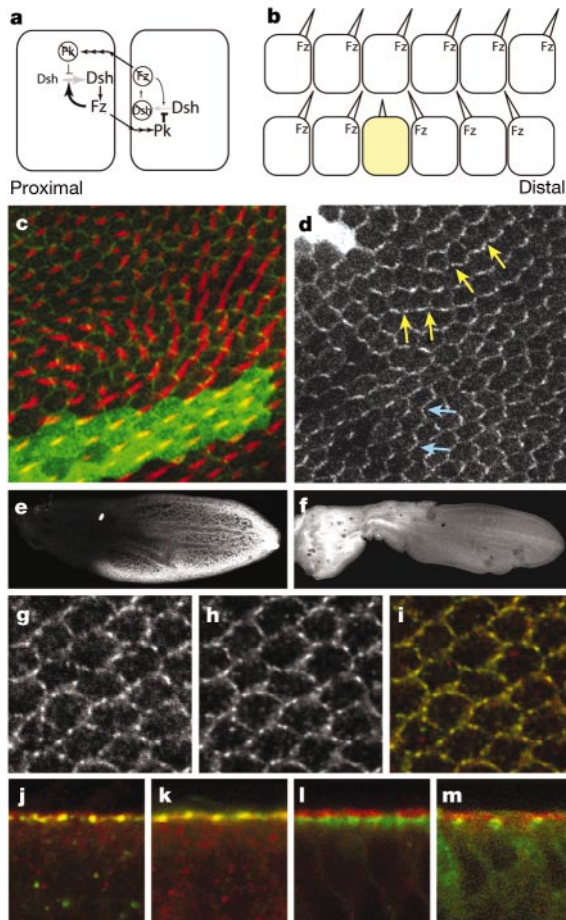


Figure 1 Four-jointed (Fj), Dachous (Ds) and Fat (Ft) in the pupal wing. In all panels, distal is to the right. **a**, The Frizzled (Fz)-dependent intercellular feedback loop⁶. Circles with a slash through them indicate protein levels that are reduced or absent; bold arrows indicate dominant interactions. **b**, Diagram showing domineering non-autonomy⁶. The top row depicts wild-type wing cells with distal Fz, and prehairs pointing distally. In the bottom row, yellow represents a cell lacking Fz. In cells distal to the *fz*^{-/-} cell, Fz localizes proximally and prehairs point proximally. **c**, *ft*^{G-RV} clone, marked by the absence of green fluorescent protein (GFP), in a 33-h pupal wing. GFP clone marker and Dsh::GFP are stained green; prehairs visualized with phalloidin are red. **d**, *ft*^{G-RV} mutant clone in a 30-h pupal wing showing GFP clone marker and Dsh::GFP. Dsh::GFP zig-zags inside the *ft*^{G-RV} clone are misaligned in the anterior portion of this image (top; yellow arrows) compared to the posterior portions (blue arrows). **e**, **f**, Anti-βGal staining of *fj-lacZ* in a 24-h pupal wing (**e**), and Ds antibody staining in a 24-h pupal wing (**f**). **g–i**, Anti-Ft (**g**), anti-Ds (**h**) and an overlay (**i**: Ft, red; Ds, green) showing co-localization at cell boundaries in a 30-h pupal wing. **j–m**, Lateral view through edge cells showing co-localization of Ft (red) and Ds (green) (**j**), co-localization of Ft (red) and Discs lost (green), a marginal zone marker (**k**), Ft (red) apical to DE-cadherin::GFP (green; adherens junctions) (**l**), and Ft (red) apical to Fz::GFP (green) (**m**). Magnifications: **c**, **d**, × 700; **e**, × 12.5; **f**, × 10; **g–i**, × 1,400; **j–m**, × 700.

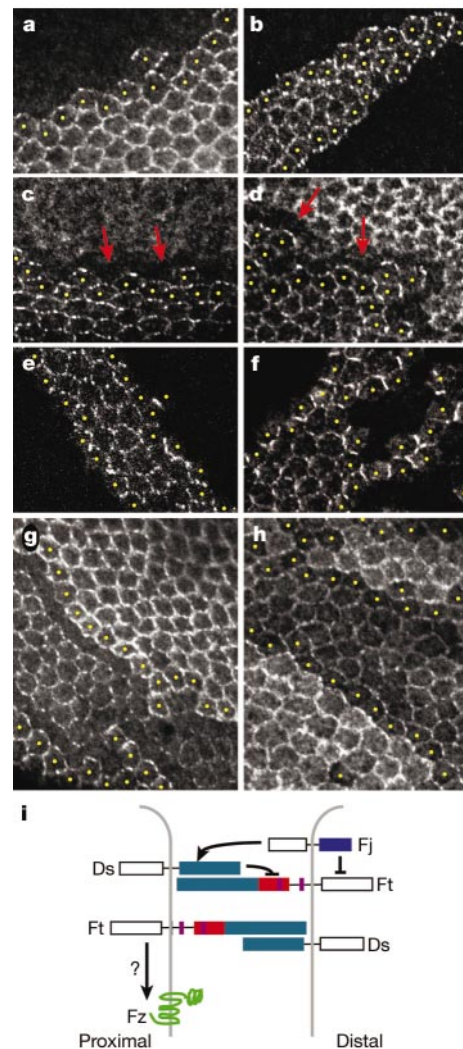


Figure 2 Dachous and Ft protein localization, and dependence on Fj. **a**, **b**, Dachous staining in a wing with *ds*^{UA071} clones (**a**), and Ft staining in a wing with *ft*^{G-RV} clones (**b**). Yellow dots mark the wild-type cells adjacent to the mutant clones in this and in the remaining panels. **c**, Dachous staining in *ft*^{G-RV} clones. **d**, Fat staining in *ds*^{UA071} clones. Note the clone boundary effect in **c** and **d** (arrows). **e**, **f**, Dachous (**e**) and Ft (**f**) staining in *ds*^{UA071} *ft*^{G-RV} double mutant clones. **g**, Dachous staining is reduced inside a *fj*^{d1} clone. Note the boundary effect. **h**, Fat staining in a wing containing a *fj*^{d1} clone is elevated. **i**, Model illustrating a possible regulatory hierarchy among Fj, Ds and Ft. Straight arrows indicate activating interactions; blunt arrows indicate antagonistic activity. Fj at the proximal side of one cell may antagonize Ft directly or indirectly through promoting Ds in the neighbouring cell. Asymmetry in Fj and/or Ds may lead to asymmetric activity of Ft, which in turn facilitates Fz activity asymmetrically, by an unknown mechanism. Magnifications: **a–h**, × 700.

throughout pupal wing development, Ds expression is highest at the hinge, and lower distally (Fig. 1f, and data not shown). These patterns are reminiscent of Ds and Fj gradients in the eye disc, where Fj expression is highest at the equator and Ds expression is highest at the poles^{19,20}. At higher resolution, Ds and Ft form punctate structures that localize together in a circumferential ring (Fig. 1g–i). However, unlike Fz and Dsh, there is no obvious asymmetry in Ds and Ft protein distribution^{15,16}. In the apical–basal axis, Ds and Ft localize together with a marginal zone marker, and are apical to the adherens junctions, where Fz and Fmi are observed (Fig. 1j–m; see also refs 10, 16).

Staining in wings bearing mutant clones reveals that both Ds and

Ft are present at the cell–cell boundary between wild-type and mutant tissues (Fig. 2a, b), suggesting that homotypic interactions are not required for their correct localization. This contrasts with Fmi, a cadherin needed for Fz localization¹⁶, which requires homotypic interactions for maintenance of its own membrane localization¹⁰. Within a *ft* mutant clone, Ds becomes diffuse, suggesting that Ds localization depends on Ft (Fig. 2c). Furthermore, Ds staining is reduced in the *ft* mutant cells immediately abutting the wild-type cells, but appears stronger at the mutant–wild-type boundary than throughout the wild-type tissue. We interpret this ‘clone-boundary effect’ to indicate that Ds protein in the first row of mutant cells selectively localizes at boundaries in contact with cells expressing Ft. Similarly, Ft staining in *ds* mutant clones shows that Ft is distributed diffusely at the cell cortex, and protein levels appear elevated (Fig. 2d). A boundary effect is again seen, suggesting that Ft selectively localizes to boundaries in contact with Ds-expressing cells. These observations support a model in which the localization of Ds and Ft to punctate marginal zone structures requires intercellular heterotypic interactions. To test this, we examined Ds and Ft distribution at the boundaries between wild-type and *ds ft* double mutant clones, and found that, indeed, both proteins are lost from the membrane (Fig. 2e, f). Therefore, Ds and Ft maintain each other’s wild-type localization in neighbouring cells.

Previous work in the eye indicated that Fj functions upstream of Ds, which is in turn upstream of Ft, and these interactions were

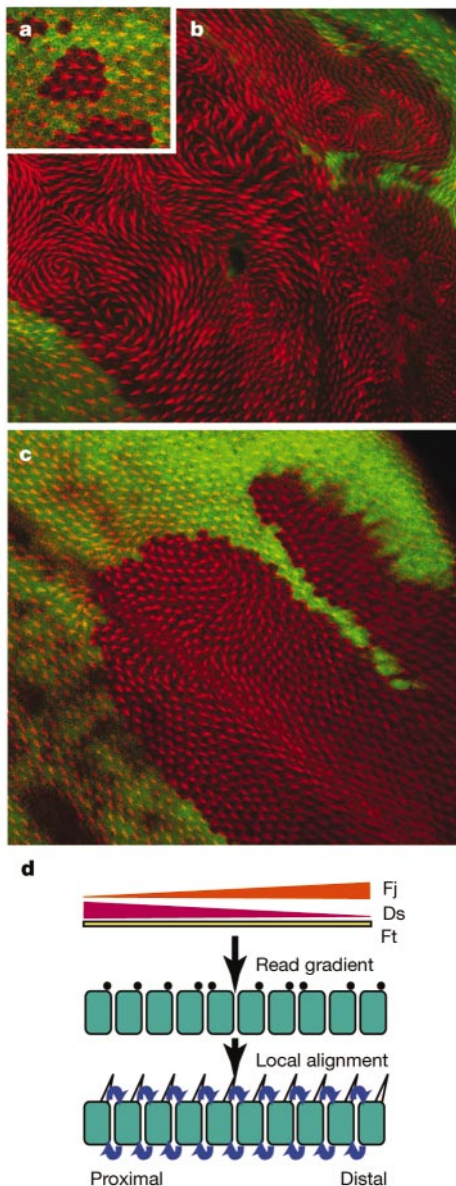


Figure 3 Fat clonal phenotypes indicate that global signalling interacts with the local alignment mechanism to achieve high fidelity. **a**, Small *ft*^{(2)td} clones show no polarity disruption. **b**, **c**, Two large *ft*^{(2)td} clones in which the prehairsts form random swirls. Distal is bottom right. **d**, Model describing the interaction between the global signalling and local alignment mechanisms. Black dots indicate the tentative position each cell places its prehair after receiving gradient input from Ft. A fraction of cells may interpret the global input erroneously. Looped arrows represent the Fz-mediated feedback loop that aligns cells locally. This alignment mechanism corrects and compensates for the cells that have initially misinterpreted the gradient input. Magnifications: **a**, $\times 400$; **b**, **c**, $\times 70$.

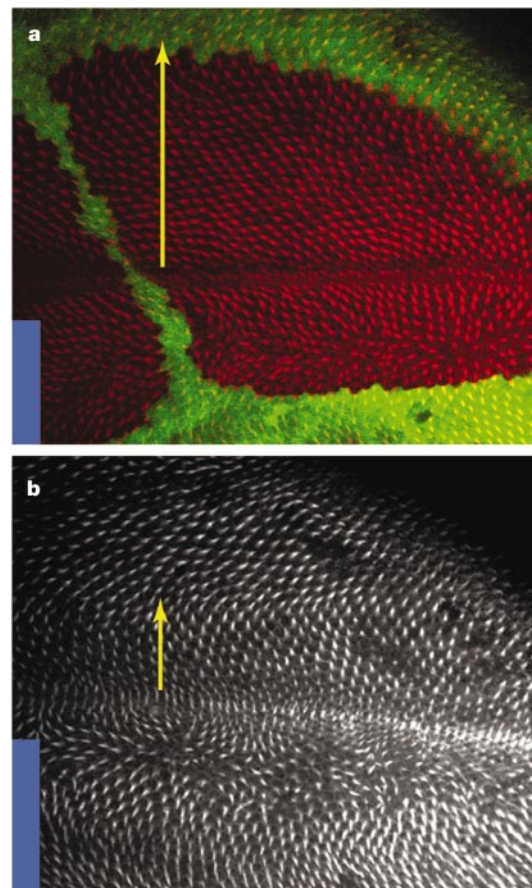


Figure 4 Loss of Ft in a large region enhances Fz domineering non-autonomy. **a**, A large *ft*^{G-IV} clone overlapping the *decapentaplegic* (*dpp*) domain where Fz is overexpressed. **b**, A wing ectopically expressing Fz in the *dpp* region in a *ft*^{G-IV} heterozygous background. The blue bars indicate the approximate extent of the *dpp* domain. Yellow arrows indicate the distance towards the anterior over which domineering non-autonomy extends. The distance is significantly greater in the *ft*^{G-IV} clone (**a**) than in the control wing (**b**). Magnifications: **a**, **b**, $\times 80$.

thought to be non-cell autonomous¹⁹. Consistent with this hierarchy, we found that Ds protein level is suppressed in *ff* wing clones (Fig. 2g). A boundary effect similar to that described above is seen, in which Ds preferentially associates with boundaries in contact with Fj-expressing cells. In contrast, Ft protein level is elevated in *ff* mutant clones (Fig. 2h). These observations support the proposal that Fj regulates Ds and Ft protein localization and activity. However, feedback regulation of *ff* expression by Ds and Ft, as seen in the eye¹⁹, or reciprocal regulation of Fj protein localization by Ds and Ft have not been ruled out.

On the basis of these results, we propose that a series of intercellular interactions transmit directional information from Fj and Ds, through Ft, to the Fz-dependent feedback loop, as shown in Fig. 2i. The expression patterns of Fj and Ds may be the source of global directional information, although it is also possible that Fj and Ds transmit directional information from another source. Although the precise relationships between these proteins are not well defined, our data indicate that a heterotypic interaction between Ds and Ft is required to influence the direction of Fz signalling.

We examined more closely the consequences of disabling this global directional signal to test the prediction that in its absence, local Fz-mediated alignment would occur, but that it would be uncoupled from the global axes of the wing. We therefore examined the phenotypes of *ft* or *ds ft* double mutant clones. These genotypes produced essentially identical results (not shown). Very large *ft* mutant clones often show marked swirls (Fig. 3b, c). Unlike *fz*, *ds*, or *ff* mutants^{23,24,26}, we observed no reproducible pattern in a given region from one wing to another. Furthermore, in these large clones, the cell–cell alignment shows continuity within and across the clone boundaries, consistent with the observation that Fz signalling is intact. In contrast, in small *ft* mutant clones, the orientation of prehair is always normal (Fig. 3a). This observation suggests the possibility that correct polarity within these small *ft* mutant clones is maintained by the intact function of the Fz-dependent feedback loop and its ability to propagate polarity. Within large *ft* clones, local alignment persists, consistent with Fz function, but the direction of this polarity signalling becomes abnormal, indicating that Fz-mediated polarity propagation cannot proceed in the correct direction over large distances in the absence of the global directional signal. We therefore propose that in the wild type, cells respond to a global directional cue mediated by Ft, but all cells may not respond accurately (Fig. 3d). Indeed, the cue is predicted to be subtle. The subsequent action of the Fz feedback loop to produce local alignment is proposed to correct errors in the initial reading of the global cue, resulting in a high-fidelity response.

Experimentally induced boundaries between cells expressing high and low Fz cause polarity to propagate in an abnormal direction through wild-type tissue. Our model predicts that this propagation will be constrained by the global directional cue mediated by Ft. For example, distal to *fz* clones, polarity is reversed in a limited domain⁶; the extent of the abnormal polarity may be limited by the global Ft-dependent signal. To test this hypothesis, we induced domineering non-autonomy by overexpressing Fz in the *decapentaplegic* (*dpp*) domain, a proximal–distal stripe between veins L3 and L4. This results in abnormal PCP both within the *dpp* domain, and at a distance anterior to the *dpp* domain, where the prehair point away from the *fz*-overexpressing cells (Fig. 4b; refs 10, 27). We then eliminated the global cue by removing *ft* in large clones that span this region, and found that the extent of domineering non-autonomy is increased, spanning the length of the clone (Fig. 4a). It has also been observed that *fz* mutant clones show enhanced domineering non-autonomy when the global signal is partially disabled in *ds* mutant wings²³. These results are consistent with our hypothesis that the Ft-mediated signal provides a directional cue that orients the action of the Fz-dependent alignment mechanism.

We have presented a model in which the interaction between the global directional signal, mediated by Ft, and a local Fz-dependent cell–cell alignment mechanism is important for achieving a high-fidelity PCP response. This interaction results in a highly robust mechanism requiring a minimum amount of complexity. How much, if any, directional information is encoded in the expression patterns of Fj and Ds, or whether they transduce information from other sources, remains to be determined. Indeed, we observe that the ability of Fj to regulate Ds and Ft localization is strongly biased towards the posterior of the wing, suggesting that an as yet unidentified component may have this function in the anterior. It will also be important to determine how Ft influences the orientation of the Fz local alignment mechanism. Our model establishes a framework for understanding the interaction between the global and local PCP signals, and how that interaction produces a high-fidelity response. □

Methods

Genetics

We used the following fly strains: *ft*^{G-rv} (ref. 5), *ft*^{(2)fd} (ref. 28), *ds*^{UAO71} *FRT40A* (ref. 23), *FRT42D* *ff*^{N7} (*fj-lacZ*; ref. 29), *FRT42D* *ff*^{d1} (ref. 24), *dsh::GFP* (ref. 15), *UASFz* (ref. 27), *dpp-Gal4* (Bloomington), *DE-cadherin::GFP*, *Fz::GFP* (ref. 16), *fz*^{R52} and *fz*^{D21} (P. Adler).

Mutant clones in the pupal wings were generated by FRT-FLP-mediated mitotic recombination³⁰. Clones were marked with ubiquitously expressed green fluorescent protein (*ubGFP*). T155 stocks were a gift from J. Duffy. The genotypes for the experiments are: *ft*^{G-rv} *FRT40A/UbGFP* *FRT40A*; *dsh::GFP/T155Gal4* *UASflp*, *ft*^{(2)fd} *FRT40A/UbGFP* *FRT40A*; *dsh::GFP/T155Gal4* *UASflp*, *ds*^{UAO71} *FRT40A/UbGFP* *FRT40A*; *dsh::GFP/T155Gal4* *UASflp*, *FRT42D* *ff*^{N7} *FRT42D* *UbGFP*; *dsh::GFP/T155Gal4* *UASflp*, *FRT42D* *ff*^{d1} *FRT42D* *UbGFP*; *T155Gal4* *UASflp* and *hsflp*; *ft*^{G-rv} *FRT40A/UbGFP* *FRT40A*; *UASFz/dppGal4*.

All flies were reared at 23–25 °C. Heat-shock clones were induced at 37 °C for 45 min. White pre-pupae were collected and allowed to age at 25 °C. All clones were induced during third instar, before PCP signalling. No perdurance of Ft or Ds was detected by immunofluorescence in clones of any size.

Immunofluorescence

Rat anti-Ft¹⁹ and anti-Ds¹⁹ (all except Fig. 1g–i) were used at 1:100. Rabbit anti-Ds (Fig. 1g, h), raised against THLPVSLPRHGHPQPRGNVGTIRM from the cytoplasmic domain, was used at 1:500. Their specificity is demonstrated by lack of staining in clones of the corresponding mutant cells (Fig. 2a, b; and data not shown). Mouse anti-βGal (Promega) was used at 1:200. Rabbit anti-Discs lost (M. Bhat and H. Bellen) was used at 1:500. Secondary Alexa488- and Alexa594-conjugated anti-rat and anti-rabbit antibodies and Alexa568-conjugated phalloidin were from Molecular Probes. Pupal wings were dissected and stained as described in ref. 15, and visualized with a Nikon laser scanning confocal microscope.

Received 27 August; accepted 12 December 2002; doi:10.1038/nature01366.

Published online 19 January 2003.

- Shulman, J. M., Perrimon, N. & Axelrod, J. D. Frizzled signaling and the developmental control of cell polarity. *Trends Genet.* **14**, 452–458 (1998).
- Adler, P. N. Planar signaling and morphogenesis in *Drosophila*. *Dev. Cell* **2**, 525–535 (2002).
- Strutt, D. I. The asymmetric subcellular localisation of components of the planar polarity pathway. *Semin. Cell Dev. Biol.* **13**, 225–231 (2002).
- Tree, D. R. P., Ma, D. & Axelrod, J. D. A three-tiered mechanism for regulation of planar cell polarity. *Semin. Cell Dev. Biol.* **13**, 217–224 (2002).
- Mahoney, P. A. et al. The fat tumour suppressor gene in *Drosophila* encodes a novel member of the cadherin gene superfamily. *Cell* **67**, 853–868 (1991).
- Vinson, C. R. & Adler, P. N. Directional non-cell autonomy and the transmission of polarity information by the *frizzled* gene of *Drosophila*. *Nature* **329**, 549–551 (1987).
- Vinson, C. R., Conover, S. & Adler, P. N. A *Drosophila* tissue polarity locus encodes a protein containing seven potential transmembrane domains. *Nature* **338**, 263–264 (1989).
- Tree, D. R. P. et al. Prickle mediates feedback amplification to generate asymmetric planar cell polarity signaling. *Cell* **109**, 371–381 (2002).
- Klingensmith, J., Nusse, R. & Perrimon, N. The *Drosophila* segment polarity gene *dishevelled* encodes a novel protein required for response to the wingless signal. *Genes Dev.* **8**, 118–130 (1994).
- Usui, T. et al. Flamingo, a seven-pass transmembrane cadherin, regulates planar cell polarity under the control of Frizzled. *Cell* **98**, 585–595 (1999).
- Chae, J. et al. The *Drosophila* tissue polarity gene *starry night* encodes a member of the protocadherin family. *Development* **126**, 5421–5429 (1999).
- Taylor, J., Abramova, N., Charlton, J. & Adler, P. N. Van gogh. A new *Drosophila* tissue polarity gene. *Genetics* **150**, 199–210 (1998).
- Wolff, T. & Rubin, G. M. Strabismus, a novel gene that regulates tissue polarity and cell fate decisions in *Drosophila*. *Development* **125**, 1149–1159 (1998).
- Gubb, D. et al. The balance between isoforms of the prickle LIM domain protein is critical for planar polarity in *Drosophila* imaginal discs. *Genes Dev.* **13**, 2315–2327 (1999).
- Axelrod, J. D. Unipolar membrane association of Dishevelled mediates Frizzled planar cell polarity signaling. *Genes Dev.* **15**, 1182–1187 (2001).
- Strutt, D. I. Asymmetric localization of Frizzled and the establishment of cell polarity in the *Drosophila* wing. *Mol. Cell* **7**, 367–375 (2001).

17. Clark, H. F. *et al.* Dachous encodes a member of the cadherin superfamily that controls imaginal disc morphogenesis in *Drosophila*. *Development* **9**, 1530–1542 (1995).
18. Villano, J. L. & Katz, F. N. Four-jointed is required for intermediate growth in the proximal-distal axis in *Drosophila*. *Development* **121**, 2767–2777 (1995).
19. Yang, C., Axelrod, J. D. & Simon, M. A. Regulation of Frizzled by Fat-like cadherins during planar polarity signaling in the *Drosophila* compound eye. *Cell* **108**, 675–688 (2002).
20. Zeidler, M. P., Perrimon, N. & Strutt, D. I. The four-jointed gene is required in the *Drosophila* eye for ommatidial polarity specification. *Curr. Biol.* **9**, 1363–1372 (1999).
21. Fanto, M. & Mlodzik, M. Asymmetric Notch activation specifies photoreceptors R3 and R4 and planar polarity in the *Drosophila* eye. *Nature* **397**, 523–526 (1999).
22. Cooper, M. T. & Bray, S. J. Frizzled regulation of Notch signalling polarizes cell fate in the *Drosophila* eye. *Nature* **397**, 526–530 (1999).
23. Adler, P. N., Charlton, J. & Liu, J. Mutations in the cadherin superfamily member gene dachous cause a tissue polarity phenotype by altering frizzled signaling. *Development* **125**, 959–968 (1998).
24. Zeidler, M. P., Perrimon, N. & Strutt, D. I. Multiple roles for four-jointed in planar polarity and limb patterning. *Dev. Biol.* **228**, 181–196 (2000).
25. Casal, J., Struhl, G. & Lawrence, P. Developmental compartments and planar polarity in *Drosophila*. *Curr. Biol.* **12**, 1189–1198 (2002).
26. Gubb, D. & Garcia-Bellido, A. A genetic analysis of the determination of cuticular polarity during development in *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* **68**, 37–57 (1982).
27. Adler, P. N., Krasnow, R. E. & Liu, J. Tissue polarity points from cells that have higher Frizzled levels towards cells that have lower Frizzled levels. *Curr. Biol.* **7**, 940–949 (1997).
28. Bryant, P. J., Huettner, B., Held, L. I. Jr, Ryser, J. & Szidonya, J. Mutations at the fat locus interfere with cell proliferation control and epithelial morphogenesis in *Drosophila*. *Dev. Biol.* **129**, 541–554 (1988).
29. Buckles, G. R., Rauskolb, C., Villano, J. L. & Katz, F. N. Four-jointed interacts with dachs, abelson and enabled and feeds back onto the Notch pathway to affect growth and segmentation in the *Drosophila* leg. *Development* **128**, 3533–3542 (2001).
30. Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).

Acknowledgements We thank D. Bilder for advice, J. Duffy and the Bloomington stock center for fly strains, and members of the Axelrod laboratory for discussions of the work and comments on the manuscript. This work was supported by the NIH (J.D.A.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.D.A. (e-mail: jaxelrod@stanford.edu).

Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia *SCL/tal-1* gene

Hanna K. A. Mikkola, Jenny Klintman, Haidi Yang, Hanno Hock, Thorsten M. Schlaeger, Yuko Fujiwara & Stuart H. Orkin

Department of Pediatric Oncology, Children's Hospital and the Dana Farber Cancer Institute, Harvard Medical School and Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

The production of blood cells is sustained throughout the lifetime of an individual by haematopoietic stem cells (HSCs)¹. Specification of HSCs from mesoderm during embryonic development requires the stem cell leukaemia *SCL/tal-1* gene product^{2–6}. Forced expression of *SCL/tal-1* strongly induces blood formation in embryos, indicating that this gene has a dominant role in commitment to haematopoiesis^{7,8}. In the adult haematopoietic system, expression of *SCL/tal-1* is enriched in HSCs and multipotent progenitors, and in erythroid and megakaryocytic lineages^{9–11}, consistent with roles for this factor in adult haematopoiesis. Here we assess by conditional gene targeting whether *SCL/tal-1* is required continuously for the identity and function of HSCs. We find that *SCL/tal-1* is dispensable for HSC engraftment, self-renewal and differentiation into myeloid and lymphoid lineages; however, the proper differentiation of erythroid

and megakaryocytic precursors is dependent on *SCL/tal-1*. Thus, *SCL/tal-1* is essential for the genesis of HSCs, but its continued expression is not essential for HSC functions. These findings contrast with lineage choice mechanisms, in which the identity of haematopoietic lineages requires continuous transcription factor expression^{12,13}.

Knockout of the *SCL/tal-1* gene is embryonic lethal in mice owing to a complete failure of haematopoiesis^{2,5}, thereby precluding a simple study of the requirement for this gene in bone marrow (BM) haematopoiesis in adults. We therefore constructed a conditional mutant¹⁴ *SCL/tal-1* strain (*SCL^{fl/fl}*; Fig. 1a). Germline Cre-recombinase-mediated deletion of the loxP-flanked (floxed) basic helix–loop–helix (bHLH) region of the gene recapitulated the phenotype of the conventional knockout, as anticipated (data not shown). To inactivate the *SCL/tal-1* gene in the adult haematopoietic compartment, we bred the *SCL^{fl/fl}* strain with MxCre mice in which Cre is expressed after the induction of interferon by administration of poly(I)–poly(C) (pIpC)¹⁵.

Whereas partial Cre-mediated deletion occurred in many tissues, excision in haematopoietic cells, including HSCs, was efficient after seven injections of pIpC administered every other day (data not shown). Peripheral blood counts were not changed significantly in pIpC-treated heterozygous (*SCL^{fl/wt}*) mice (data not shown). In pIpC-treated *SCL^{fl/fl}* mice, but not in control *SCL^{fl/wt}* mice, excision of *SCL/tal-1* resulted in moderate anaemia and thrombocytopenia, with nadirs at 4 and 2 weeks after the initiation of pIpC treatment, respectively, and in about 5% lethality (data not shown). The floxed allele was excised completely in haematopoietic cells of pIpC-treated *SCL^{fl/wt}* mice, as assessed by three-primer, genomic polymerase chain reaction (PCR) of the DNA from peripheral blood cells and individual colonies grown from BM cells (Fig. 1b and data not shown). In surviving pIpC-treated *SCL^{fl/fl}* mice, however, an intact floxed allele was detected by PCR in a fraction of peripheral blood and BM cells (Fig. 1b and data not shown). These data indicate that *SCL/tal-1* may be required, as thought previously, for

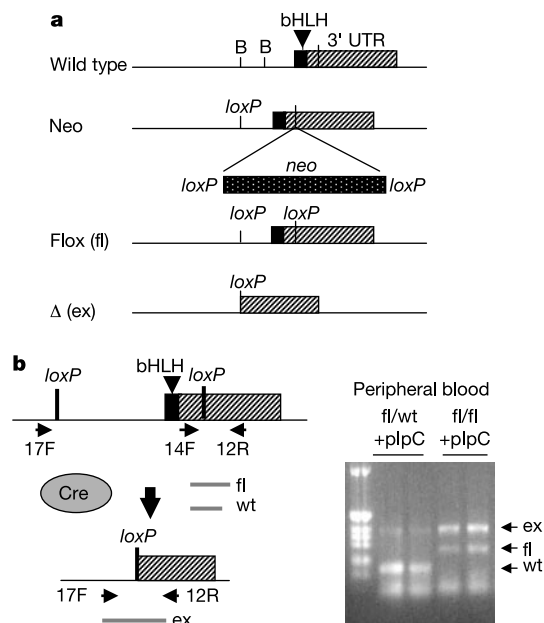


Figure 1 Inducible inactivation of the *SCL/tal-1* locus *in vivo*. **a**, Conditional targeting of the *SCL/tal-1* locus. The floxed neomycin cassette was introduced at a unique *Xba*I site in the 3' untranslated region (3' UTR). B, *Bam*HI site. **b**, Deletion of the floxed region was monitored by semiquantitative, three-primer PCR. Peripheral blood PCR from two representative examples at 12 weeks after pIpC administration is shown on the right. ex, excised; fl, floxed; wt, wild-type products.

the survival of treated mice, but they do not define the nature of the deficits in the absence of *SCL/tal-1*.

To examine the consequences of *SCL/tal-1* gene inactivation on haematopoiesis under nonselective conditions, we generated mice with chimaeric BM cells in which conditionally targeted cells could be distinguished from support cells by surface marker expression. *SCL^{fl/fl}* or *SCL^{fl/wt}* total BM cells carrying the CD45.2 allele of the pan-haematopoietic marker CD45 were transplanted together with CD45.1⁺ support BM cells into CD45.1⁺/CD45.2⁺ heterozygous recipients (Fig. 2a). In addition to competing with excised cells, support BM cells ensure the production of all blood cell lineages, thereby facilitating an extended study of haematopoiesis and obviating the selection for or against cells lacking *SCL/tal-1* expression. Half the mice in each group were treated with pIpC to induce excision of the floxed allele. The contribution of excised cells to blood and haematopoietic organs was monitored by surface expression of the CD45.2 allele (Fig. 2b) and genomic PCR (Fig. 2e).

Notably, *SCL^{fl/fl}* and *SCL^{fl/wt}* CD45.2⁺ cells contributed compar-

ably to peripheral blood cells in recipient mice for more than 6 months in spite of excision of the *SCL* locus (Fig. 2c). Excised *SCL^{fl/fl}* CD45.2⁺ cells were represented in Gr1⁺ and Mac1⁺ myeloid cells (granulocytes and macrophages, respectively), CD3⁺, CD4⁺ and CD8⁺ T-lymphoid cells, B220⁺ and CD19⁺ B-lymphoid cells and DX5⁺ NK cells (Fig. 2d). Excision of the gene-targeted allele in *SCL^{fl/fl}* and *SCL^{fl/wt}* CD45.2⁺ peripheral white blood cells was quantitative in mice previously treated with pIpC (Fig. 2e). Thus, continuous expression of *SCL/tal-1* is not required for maintaining the myeloid (granulocyte and macrophage) and lymphoid lineages during steady-state haematopoiesis. However, we have observed a slight predominance of T cells relative to B cells in peripheral blood lymphocytes after excision of *SCL*, which suggests that *SCL* expression may influence lymphoid development.

The differentiation potential of *SCL/tal-1*[−] cells was studied further by *in vitro* colony assays of CD45.2⁺ BM cells purified by fluorescence-activated cell sorting (FACS; Fig. 3a). The overall frequency of colony formation was not altered appreciably in *SCL^{fl/fl}* or *SCL^{fl/wt}* cells despite Cre-mediated inactivation of the *SCL* gene, as detected by single-colony PCR (Fig. 3a) and PCR with reverse transcription (RT-PCR) from pooled colonies (data not shown). In agreement with *in vivo* findings, myeloid differentiation was unaffected: granulocyte and macrophage cells were present in granulocyte/macrophage colonies derived from Cre-deleted *SCL^{fl/fl}* and *SCL^{fl/wt}* CD45.2⁺ BM cells (Fig. 3b). By contrast, erythroid and megakaryocytic cells were not identified in cultures of *SCL^{fl/fl}* CD45.2⁺ cells from pIpC-treated mice (Fig. 3b). Lack of histochemical staining both for acetylcholinesterase activity (for megakaryocytes) and with benzidine reagent (for erythroid cells)

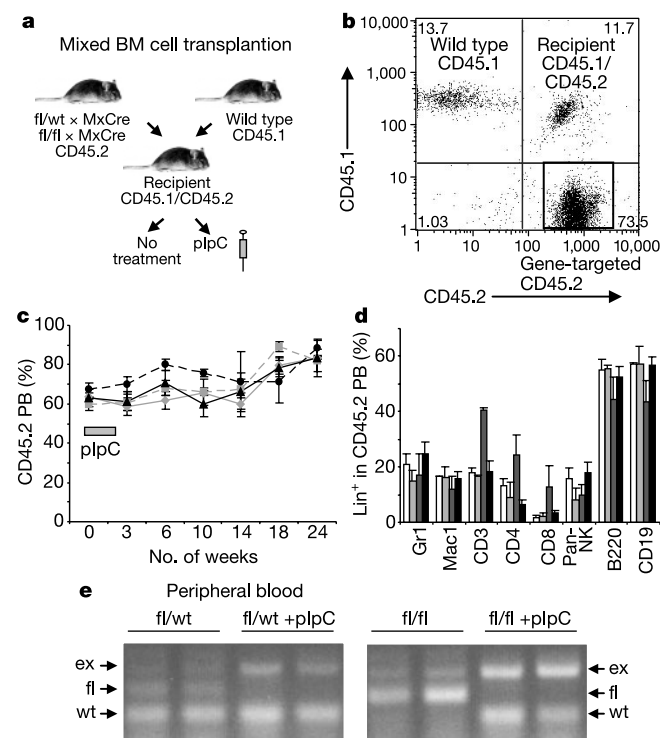


Figure 2 Continuing lymphopoiesis and myeloopoiesis in the absence of *SCL/tal-1*. **a**, Mice with chimaeric bone marrow (BM) cells were generated by transplanting a 2:1 ratio of conditionally targeted BM (CD45.2⁺) cells with support BM (CD45.1⁺) cells into irradiated recipients (CD45.1⁺/CD45.2⁺). Half of the mice were subjected to pIpC induction, resulting in loss of *SCL* in CD45.2⁺ haematopoietic cells. **b**, Contribution from the conditionally targeted BM cells in peripheral blood was monitored by surface expression of CD45.2 antigen. **c**, Contribution of CD45.2⁺ cells in peripheral blood during 6 months follow up ($n = 4-6$ mice per group). Unbroken grey lines, pIpC-treated *SCL^{fl/wt}* mice; broken grey lines, untreated *SCL^{fl/wt}* mice; unbroken black lines, pIpC-treated *SCL^{fl/fl}* mice; broken black lines, untreated *SCL^{fl/fl}* mice. The period of pIpC treatment is indicated by the shaded box. **d**, Presence of Gr1⁺ and Mac1⁺ myeloid cells CD3⁺, CD4⁺ and CD8⁺ T cells, B220⁺ and CD19⁺ B cells, and DX5⁺ NK cells in CD45.2⁺ peripheral blood cells at 6 months after pIpC treatment ($n = 3$ per group). White bars, pIpC-treated *SCL^{fl/wt}* mice; light grey bars, untreated *SCL^{fl/wt}* mice; dark grey bars, pIpC-treated *SCL^{fl/fl}* mice; black bars, untreated *SCL^{fl/fl}* mice. **e**, PCR shows excision of the floxed allele followed by pIpC induction throughout the study. Representative examples at 6 weeks after pIpC treatment are shown. The faint presence of an excised band in the absence of pIpC treatment reflects low-level spontaneous induction of the Cre transgene in MxCre mice¹⁴.

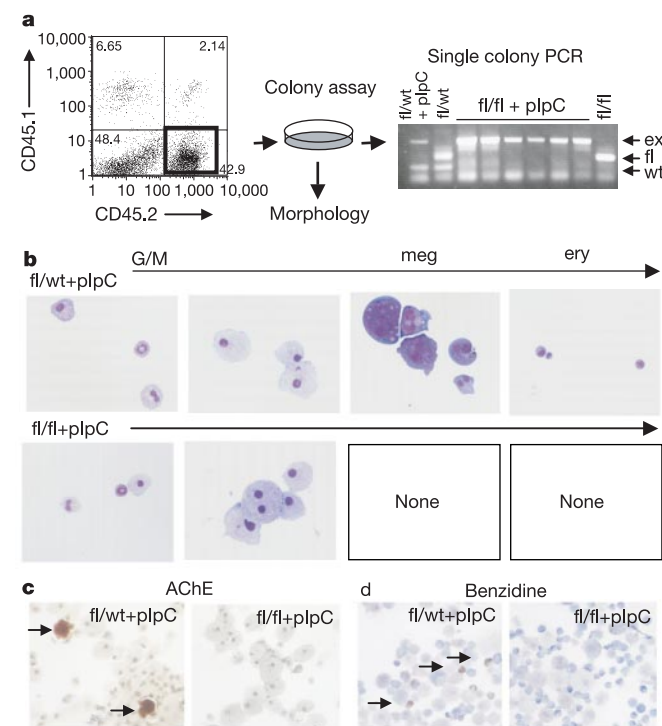


Figure 3 Requirement for *SCL* in erythroid and megakaryocytic differentiation *in vitro*. **a**, FACS-purified CD45.2⁺ BM cells from transplant recipients treated 2 months earlier with pIpC were cultured in methylcellulose. Excision of the floxed fragment in individual colonies was verified by single-colony PCR (representative examples are shown). **b-d**, *In vitro* differentiation of CD45.2⁺ *SCL*-deficient BM cells (*fl/fl* + pIpC) shows absence of cells with erythroid (ery) and megakaryocyte (meg) morphology (**b**), acetylcholinesterase activity (**c**) and benzidine staining (**d**). For acetylcholinesterase and benzidine assays, colonies were pooled from whole plates cultured with a mix of haematopoietic growth factors.

confirmed the absence of these cells (Fig. 3c, d). These findings provide evidence that *SCL/tal-1* expression, although dispensable for myeloid and lymphoid haematopoiesis, is required selectively for proper erythroid and megakaryocytic development.

To define more precisely the stage at which *SCL/tal-1* becomes necessary in the erythroid lineage, we analysed BM cells from transplanted mice for erythroid differentiation markers. In cells from *SCL^{fl/wt}* mice treated with pIpC, a normal pattern of erythro-

blast development through transferrin receptor ($CD71$)⁺ Ter119⁺ cells was observed (Fig. 4a, left, upper boxed region)¹⁶. The FACS profile of BM cells of pIpC-treated *SCL^{fl/fl}* mice showed an abnormal Ter119^{lo/-}CD71⁺ population (Fig. 4a, right, left boxed region). PCR analysis of the Ter119⁻CD71⁺ and Ter119⁺CD71⁺ populations showed that the normal erythroblast population in *SCL^{fl/fl}* mice was derived from cells with a wild-type *SCL/tal-1* genotype (that is, CD45.1⁺ support cells), whereas the abnormal population contained cells with a deleted locus (Fig. 4b). In addition, FACS-sorted Ter119^{lo/-}CD71⁺ cells, which appeared basophilic, failed to react with benzidine reagent (Fig. 4c, d). The Ter119^{lo/-}CD71⁺

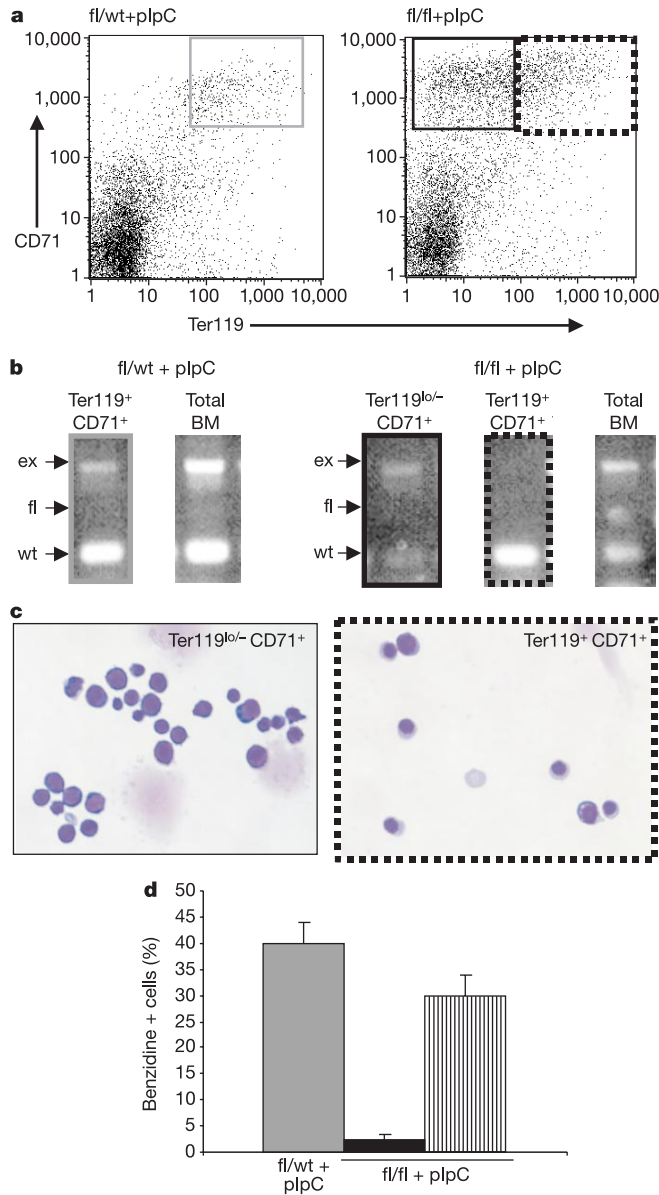


Figure 4 Requirement for *SCL* in erythroblast maturation in the bone marrow. **a**, FACS analysis of BM cells from transplanted, pIpC-treated mice shows an accumulation of abnormal Ter119^{lo/-}CD71⁺ cell population in mice reconstituted with *SCL^{fl/fl}* cells and treated with pIpC (solid black box). **b**, Semiquantitative PCR of FACS-purified BM fractions in pIpC-treated *SCL^{fl/fl}* mice shows contribution of *SCL*-deficient cells (ex) in the Ter119^{lo/-}CD71⁺ population (solid black box in **a**), whereas Ter119⁺CD71⁺ erythroblasts are derived from the support BM (wt) cells (broken black box in **a**). **c**, May-Grunwald-Giemsa staining of sorted subfractions from chimaeric BM cells. Left, Ter119^{lo/-}CD71⁺ cells from chimaeric *SCL^{fl/fl}* BM cells; right, Ter119⁺CD71⁺ cells from chimaeric *SCL^{fl/fl}* BM cells. **d**, Benzidine staining of sorted subfractions from chimaeric BM cells. Grey bar, Ter119⁺CD71⁺ cells from chimaeric *SCL^{fl/wt}* BM cells; filled black bar, Ter119^{lo/-}CD71⁺ cells from chimaeric *SCL^{fl/fl}* BM cells; striped bar, Ter119⁺CD71⁺ cells from chimaeric *SCL^{fl/fl}* BM cells.

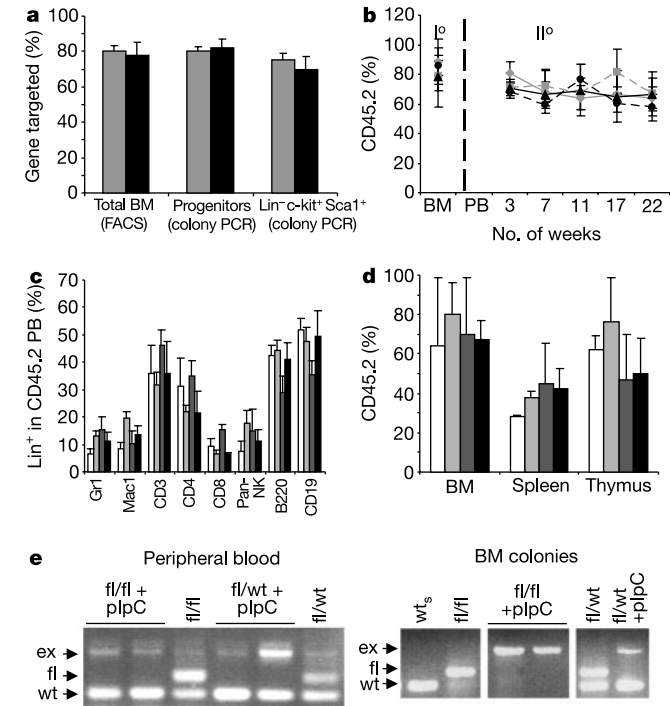


Figure 5 Haematopoietic stem cell function in the absence of *SCL/tal-1*. The effect of *SCL* loss on maintenance and function of HSCs was studied by serial BM transplantation. **a**, Contribution of the gene-targeted cells to total BM cells, and progenitor and Lin⁻c-kit⁺Sca1⁺ subsets in primary recipients was studied 4 months after pIpC treatment. Total BM is the contribution of CD45.2⁺ cells to unfractionated BM cells by FACS analysis; progenitors is the percentage of gene-targeted colonies grown from total BM cells in single-colony PCR; Lin⁻c-kit⁺Sca1⁺ is the percentage of gene-targeted colonies grown from the Lin⁻c-kit⁺Sca1⁺ fraction in single-colony PCR; $n = 40-60$ colonies per group. Grey bars, pIpC-treated *SCL^{fl/wt}* mice; black bars, pIpC-treated *SCL^{fl/fl}* mice. All colonies derived from gene-targeted cells from pIpC-treated mice showed excision of the *SCL* locus. **b**, Chimaeric BM cells from primary recipients reconstituted with a 2:1 ratio of gene-targeted and support BM cells were transplanted into secondary recipients, and the contribution of the CD45.2⁺ cells in peripheral blood after transplantation was determined by FACS. Shown are the means $\pm$ s.e.m. ($n = 4-8$ per group). Unbroken grey lines, pIpC-treated *SCL^{fl/wt}* mice; broken grey lines, untreated *SCL^{fl/wt}* mice; unbroken black lines, pIpC-treated *SCL^{fl/fl}* mice; broken black lines, untreated *SCL^{fl/fl}* mice. **c**, Multilineage engraftment was evident for more than 6 months after secondary transplantation ($n = 3$ per group). **d**, Contribution of transplanted CD45.2⁺ cells in haematopoietic organs was documented 2 months after secondary transplantation ($n = 2$ per group). In **c** and **d**, white bars, pIpC-treated *SCL^{fl/wt}* mice; light grey bars, untreated *SCL^{fl/wt}* mice; dark grey bars, pIpC-treated *SCL^{fl/fl}* mice; black bars, untreated *SCL^{fl/fl}* mice. **e**, Semiquantitative PCR shows quantitative excision of floxed alleles in gene-targeted peripheral blood cells (left) and BM colonies (right) from mice subjected to pIpC treatment. Representative examples at 6 and 2 months after secondary transplantation are shown. wt_s indicates a colony derived from support BM cells. The faint excised band observed in the absence of pIpC treatment reflects low-level spontaneous induction of the Cre transgene in MxCre mice¹⁴.

population encompasses erythroid precursors that are unable to complete cellular maturation in the absence of *SCL/tal-1*.

Persistent myeloid and lymphoid haematopoiesis after inactivation of the *SCL/tal-1* gene could not have been anticipated by earlier findings in the context of embryonic development. Although unlikely, the above findings did not eliminate the possibility that haematopoiesis under these experimental conditions might be sustained by *SCL/tal-1*⁻ committed haematopoietic progenitors rather than by *SCL/tal-1*⁻ HSCs. In the chimaeric BM cells of pIpC-treated transplanted mice, the proportions of excised *SCL*^{fl/fl} and *SCL*^{fl/wt} CD45.2⁺ cells in the Lin⁻c-kit⁺Sca1⁺ compartment, clonogenic progenitors and unfractionated BM cells were comparable at 4 months, as analysed by FACS sorting and single-colony PCR (Fig. 5a). Thus, *SCL/tal-1*⁻ HSCs or progenitors are not apparently disadvantaged during steady-state haematopoiesis.

To address the requirement, if any, for *SCL/tal-1* in HSCs more directly, we serially transplanted BM cells from pIpC-treated primary recipients into secondary recipients. Excised CD45.2⁺ cells contributed comparably to peripheral blood and haematopoietic organs after secondary transplantation (Fig. 5b, d). Multilineage engraftment was observed for at least 5 months, confirming the lack of selection against *SCL/tal-1*⁻ HSCs relative to *SCL/tal-1*⁺ HSCs *in vivo* (Fig. 5c). PCR analysis of DNA from peripheral blood leukocytes and BM-derived colonies (*n* = 20–30 colonies per group) showed that haematopoiesis in secondary recipients was directed by *SCL*^{fl/fl} cells, in which excision of the floxed region had indeed taken place (Fig. 5e). Thus, despite expression of *SCL/tal-1* in normal HSCs, loss of *SCL/tal-1* does not seem to impair considerably HSC properties, including engraftment, self-renewal and multipotency.

Although *SCL/tal-1* is a member of the bHLH family of transcription factors, its DNA-binding activity is dispensable for haematopoietic commitment⁴. DNA binding is necessary, however, for proper maturation of erythroid and megakaryocytic precursors⁴. Notably, this distinction is paralleled here by a continuous requirement for *SCL/tal-1* expression for normal differentiation of these two lineages and its dispensability for maintaining HSC properties. Studies addressing mechanisms of lineage choice in the haematopoietic system have suggested that transcription factors, acting in concert and antagonistically, program individual lineages¹² and that continuous expression of essential factors is necessary to maintain the identity of committed cells¹³. Our findings provide an alternative perspective on the *in vivo* role of the essential stem cell factor, *SCL/tal-1*, in haematopoietic development.

SCL/tal-1, together with its partner protein LMO2 (refs 17, 18), directs haematopoietic fate commitment during embryogenesis. But in the adult, continuous expression of *SCL/tal-1* is not required for maintaining HSCs or for repopulating the haematopoietic system. In its absence, HSCs neither fail to survive nor lose HSC identity or 'stemness', indicating that the transition from mesoderm (or haemangioblasts)^{19–21} to HSCs is not reversed on loss of *SCL*. Results of single haematopoietic colony PCR analyses (for example, see Fig. 5e) exclude the fusion of support BM cells contributing a normal *SCL/tal-1* allele with excised *SCL*^{fl/fl} cells as an explanation for sustained haematopoiesis^{22,23}. The apparent stability of the HSC phenotype in the absence of *SCL/tal-1* may reflect complex epigenetic regulatory circuits established in a context-dependent manner during embryogenesis. Our results now distinguish two classes of haematopoietic 'stem cell' transcription factors: those required for HSC genesis, and those required for later HSC properties such as long-term repopulating activity and multipotency. □

Methods

Conditional gene targeting

The *SCL/tal-1* locus was modified by conditional gene targeting¹⁴ in embryonic stem (ES) cells. The targeting construct introduced an upstream *loxP* site in the intron preceding the bHLH region and a floxed *neomycin* resistance cassette in the 3' untranslated region. The

neomycin cassette was removed subsequently from targeted ES cell clones by transient expression of Cre, yielding the final floxed locus. Details are available from S.H.O. on request. The interferon inducible mouse strain (MxCre)¹⁵ was used to inactivate the *SCL/tal-1* gene in adult mice. Intraperitoneal injections of pIpC (seven injections of 20 µg per g (body) weight every other day for 14 d) result in endogenous secretion of α- and β-interferon, activate the Mx promoter and yield abundant expression of Cre in haematopoietic cells.

Semiquantitative PCR

The contribution of intact (fl) or excised (ex) gene-targeted cells and support cells (wt) was monitored by semiquantitative three-primer PCR (forward for 'fl' or 'wt' alleles: 14F, 5'-TGAAGATGGCAGCTCTTCTC-3'; forward for excised allele: 17F, 5'-ACCAACAACAACCGGGTGAAG-3'; reverse for all alleles: 12R, 5'-ATGGGAGAAAGGCAAGGCAG-3').

BM transplantation

We transplanted 1 × 10⁶ unfractionated BM cells from MxCre/SCL conditional (*SCL*^{fl/wt} and *SCL*^{fl/fl}, *n* = 4 per genotype) compound mice (CD45.2⁺) together with 0.5 × 10⁶ support BM cells from congenic wild-type C57BL/6-Ly5.1 (CD45.1⁺) mice (Jackson Labs) into irradiated (1,300 rad) recipients that were heterozygous for CD45.2 and CD45.1 alleles (*n* = 18 per genotype). Contribution from the conditionally targeted cells was analysed first at 4 weeks after transplantation; thereafter, half of the *SCL*^{fl/wt} and *SCL*^{fl/fl} recipients were induced with pIpC, resulting in excision of the conditionally targeted allele/alleles. Secondary transplantation of chimaeric BM cells was done by transplanting 5 × 10⁶ unfractionated BM cells of the primary recipients into irradiated CD45.1⁺/CD45.2⁺ secondary recipients (*n* = 10 per group).

FACS analysis and sorting

The contribution of gene-targeted CD45.2⁺ haematopoietic cells in peripheral blood and haematopoietic organs (BM, spleen and thymus) was verified by staining with biotin-conjugated antibody against CD45.2 (BD Pharmingen) and FITC-conjugated antibody against CD45.1, followed by secondary staining with allophycocyanin-streptavidin. Lineage analysis was done by co-staining with phycoerythrin (PE)-conjugated antibodies against CD3, CD4, CD8, B220, CD19, Gr1, Mac1 or DX5 (Pan-NK). Erythroid differentiation in the BM was analysed by staining with Ter119-biotin/streptavidin-FITC and CD71-PE. Lin⁻c-kit⁺sca1⁺ BM cells were sorted after staining with c-kit-PE, sca1-FITC and Lin-biotin/streptavidin-APC (Lin⁺ cells are CD3⁺, B220⁺, Gr1⁺, Mac1⁺ and Ter119⁺).

Methylcellulose culture

Unfractionated or FACS-sorted CD45.2⁺ BM was cultured in methylcellulose (#3234, Stem Cell Technologies) supplemented with 50 ng ml⁻¹ stem cell factor, 10 ng ml⁻¹ interleukin 3 (IL-3), 10 ng ml⁻¹ IL-6, 5 ng ml⁻¹ Tpo and 2 U ml⁻¹ Epo. We scored colonies by microscopy, single-colony PCR and May-Grunwald-Giemsa staining. For single-colony PCR, individual colonies were picked and lysed in 2 × PCR buffer (20 mM Tris-HCl, 3 mM MgCl₂ and 100 mM KCl), 0.45% Nonidet P-40, 0.45% Tween-20 and 100 µg ml⁻¹ proteinase K for 2 h at 55 °C followed by 15 min of inactivation at 95 °C. Erythroid and megakaryocytic differentiation in the cultures was evaluated by benzidine reagent positivity and acetylcholinesterase histochemistry²⁴.

Received 27 September; accepted 26 November 2002; doi:10.1038/nature01345.

Published online 19 January 2003.

- Weissman, I. L. Translating stem and progenitor cell biology to the clinic: barriers and opportunities. *Science* **287**, 1442–1446 (2000).
- Shivdasani, R. A., Mayer, E. L. & Orkin, S. H. Absence of blood formation in mice lacking the T-cell leukaemia oncogene *tal-1*/SCL. *Nature* **373**, 432–443 (1995).
- Porcher, C. *et al.* The T cell leukemia oncogene *SCL/tal-1* is essential for development of all hematopoietic lineages. *Cell* **86**, 47–57 (1996).
- Porcher, C., Liao, E. C., Fujiwara, Y., Zon, L. I. & Orkin, S. H. Specification of hematopoietic and vascular development by the bHLH transcription factor SCL without direct DNA binding. *Development* **126**, 4603–4615 (1999).
- Robb, L. *et al.* Absence of yolk sac hematopoiesis from mice with a targeted disruption of the *scl* gene. *Proc. Natl Acad. Sci. USA* **92**, 7075–7079 (1995).
- Robb, L. *et al.* The *scl* gene product is required for the generation of all hematopoietic lineages in the adult mouse. *EMBO J.* **15**, 4123–4129 (1996).
- Gering, M., Rodaway, A. R., Gottgens, B., Patient, R. K. & Green, A. R. The SCL gene specifies haemangioblast development from early mesoderm. *EMBO J.* **17**, 4029–4045 (1998).
- Mead, P. E., Deconinck, A. E., Huber, T. L., Orkin, S. H. & Zon, L. I. Primitive erythropoiesis in the *Xenopus* embryo: the synergistic role of LMO-2, SCL and GATA-binding proteins. *Development* **128**, 2301–2308 (2001).
- Elefanti, A. G. *et al.* Characterization of hematopoietic progenitor cells that express the transcription factor SCL, using a lacZ 'knock-in' strategy. *Proc. Natl Acad. Sci. USA* **95**, 11897–11902 (1998).
- Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000).
- Ramallo-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C. & Melton, D. A. 'Stemness': transcriptional profiling of embryonic and adult stem cells. *Science* **298**, 597–600 (2002).
- Orkin, S. H. Diversification of haematopoietic stem cells to specific lineages. *Nature Rev. Genet.* **1**, 57–64 (2000).
- Mikkola, I., Heavey, B., Horcher, M. & Busslinger, M. Reversion of B cell commitment upon loss of Pax5 expression. *Science* **297**, 110–113 (2002).
- Gu, H., Marth, J. D., Orban, P. C., Mossman, H. & Rajewsky, K. Deletion of a DNA polymerase-β gene segment in T cells using cell type-specific gene targeting. *Science* **265**, 103–106 (1994).

15. Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
16. Socolovsky, M. *et al.* Ineffective erythropoiesis in *Stat5a*^{-/-}*Stat5b*^{-/-} mice due to decreased survival of early erythroblasts. *Blood* **98**, 3261–3273 (2001).
17. Wadman, I. A. *et al.* The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* **16**, 3145–3157 (1997).
18. Yamada, Y. *et al.* The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proc. Natl Acad. Sci. USA* **95**, 3890–3895 (1998).
19. Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C. & Keller, G. A common precursor for hematopoietic and endothelial cells. *Development* **125**, 725–732 (1998).
20. Robertson, S. M., Kennedy, M., Shannon, J. M. & Keller, G. A transitional stage in the commitment of mesoderm to hematopoiesis requiring the transcription factor *SCL/tal-1*. *Development* **127**, 2447–2459 (2000).
21. Faloony, P. *et al.* Basic fibroblast growth factor positively regulates hematopoietic development. *Development* **127**, 1931–1941 (2000).
22. Terada, N. *et al.* Bone marrow cells adopt the phenotype of other cells by spontaneous cell fusion. *Nature* **416**, 542–545 (2002).
23. Ying, Q. L., Nichols, J., Evans, E. P. & Smith, A. G. Changing potency by spontaneous fusion. *Nature* **416**, 545–548 (2002).
24. Cantor, A. B., Katz, S. G. & Orkin, S. H. Distinct domains of the GATA-1 cofactor FOG-1 differentially influence erythroid versus megakaryocytic maturation. *Mol. Cell. Biol.* **22**, 4268–4279 (2002).

Acknowledgements We thank D. Traver for discussions; J. Dailey and S. Lazo-Kallanian for cell sorting; K. Rajewsky for MxCre mice; and A. Williams and S. Galusha for assistance in generating the conditional *SCL/tal-1* strain. H.K.A.M. received support from the Finnish Cultural Foundation and the Academy of Finland. This work was supported in part by a grant from the NIH to S.H.O., who is an Investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.H.O. (e-mail: stuart_orkin@dfci.harvard.edu).

Crystal structure of the human angiotensin-converting enzyme–lisinopril complex

Ramanathan Natesh*, Sylva L. U. Schwager†, Edward D. Sturrock† & K. Ravi Acharya*

* Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK

† Division of Medical Biochemistry and MRC/UCT Liver Research Centre, University of Cape Town Medical School, Observatory 7925, South Africa

Angiotensin-converting enzyme (ACE) has a critical role in cardiovascular function by cleaving the carboxy terminal His-Leu dipeptide from angiotensin I to produce a potent vasopressor octapeptide, angiotensin II. Inhibitors of ACE are a first line of therapy for hypertension, heart failure, myocardial infarction and diabetic nephropathy. Notably, these inhibitors were developed without knowledge of the structure of human ACE, but were instead designed on the basis of an assumed mechanistic homology with carboxypeptidase A¹. Here we present the X-ray structure of human testicular ACE and its complex with one of the most widely used inhibitors, lisinopril (*N*²-[(S)-1-carboxy-3-phenylpropyl]-L-lysyl-L-proline; also known as Prinivil or Zestril), at 2.0 Å resolution. Analysis of the three-dimensional structure of ACE shows that it bears little similarity to that of carboxypeptidase A, but instead resembles neurolysin² and *Pyrococcus furiosus* carboxypeptidase³—zinc metalloproteinases with no detectable sequence similarity to ACE. The structure provides an opportunity to design domain-selective ACE inhibitors that may exhibit new pharmacological profiles.

Angiotensin-converting enzyme (also known as peptidyl dipeptidase A, EC 3.4.15.1 or ACE) is a type-I membrane-anchored dipeptidyl carboxypeptidase that is essential for blood pressure regulation and electrolyte homeostasis through the renin–angiotensin–aldosterone system. There are two isoforms of ACE that are transcribed from the same gene in a tissue-specific manner. In somatic tissues it exists as a glycoprotein composed of a single, large polypeptide chain of 1,277 amino acids, whereas in sperm cells it is a lower-molecular-mass glycoform of 701 amino acids. The somatic form consists of two homologous domains (N and C domain), each of which contains an active site with a conserved HEXXH zinc-binding motif⁴, where the two histidines are zinc ligands, with a glutamate 24 residues downstream forming the third ligand⁵. The two domains differ in their substrate specificities, inhibitor and chloride activation profiles, and physiological functions⁶. There are two N-domain-specific substrates: the peptide *N*-acetyl-seryl-aspartyl-lysyl-proline, which regulates haematopoietic stem cell differentiation and proliferation; and the bradykinin-potentiating peptide angiotensin-(1-7)⁷. On the other hand, the active sites of both domains catalyse the hydrolysis of angiotensin I and the vasodilator bradykinin with similar efficiency. However, inhibition of the N domain with a phosphinic peptide RXP407 has no effect on blood pressure regulation⁷, and expression in transgenic mice of the N domain alone produces a phenotype similar to that seen in complete ACE knockout mice⁸. Thus, the C domain seems to be necessary and sufficient for controlling blood pressure and cardiovascular function, suggesting that the C domain is the dominant angiotensin-converting site.

Testis ACE is identical to the C-terminal half of somatic ACE, except for a unique 36-residue sequence constituting its amino terminus⁹. We used a truncated version, tACEΔ36NJ (a mutant with full enzymatic activity that is truncated at Ser 625 and that lacks the O-glycan-rich, N-terminal 36 residues and hydrophobic transmembrane domain), for structure determination. The carbohydrates were modified to minimize oligosaccharide-based heterogeneity¹⁰. We refer to this as tACE.

The structure of tACE (residues 37–625) adopts an overall ellipsoid shape (dimensions approximately 72 × 57 × 48 Å) with a central groove that extends for about 30 Å into the molecule and divides the protein into two ‘subdomains’ (labelled I and II in Fig. 1a). The boundaries of the groove are provided by helices α13, α14, α15, α17 and strand β4. The structure of the tACE–lisinopril complex was used to locate the active site of the molecule. On top of the molecule there is an N-terminal ‘lid’ formed by helices α1, α2 and α3 (all three containing several charged residues), which seems to restrict the access of large polypeptides to the active-site cleft and thereby accounts for the enzyme’s inability to hydrolyse large, folded substrates (Fig. 1b). The structure of tACE is predominantly helical with 27 helices (almost equally distributed in both subdomains), 20 α-helices and seven ₃₁₀-helices (Fig. 1a, c). The only β-structure, accounting for 4% of all residues, occurs as six relatively short strands, two of which are located near the active site (Fig. 1a). The long central helix α15 packs diagonally across the two kinked helices α6 and α8 (both of them contain a proline residue).

Residues Asp 40 (α1) and Gly 615 (five residues downstream of α20) define the N and C termini of the ectodomain, respectively, and this is in agreement with previous tACE mutagenesis and cleavage-secretion studies¹¹. Residues Ser 435 to Gly 438 are disordered in the structure. A total of 504 water molecules were identified in the native structure. The active-site pocket is occupied by several ordered water molecules. All six glycosylation sites (g1–6, Fig. 1c) are located on the surface of the tACE molecule. Weak electron density was observed for *N*-linked carbohydrates in all of the glycosylation sites and modelled with an *N*-acetylglucosamine moiety.

Zinc is an important catalytic component of ACE^{6,12}. One highly ordered zinc ion (*B*-factor = 16.3 Å² in tACE-native) is bound at

the active site. Helix $\alpha 13$ contains the HEXXH zinc-binding motif, with its two zinc-coordinating histidines (His 383 and His 387). Additional coordination is provided by Glu 411 on helix $\alpha 14$ and an acetate ion (from the crystallization medium). The role of a zinc ion in ACE catalysis was thought to be analogous to that in thermolysin⁴, and our structural data confirm that the zinc-binding sites in both proteins are indeed very similar (root mean square (r.m.s.) deviation of 0.52 Å (all atoms) for residues that are zinc ligands in both proteins), except for the acetate ion that is replaced by a water molecule in the coordination sphere in thermolysin.

Substrate hydrolysis by ACE is activated by chloride ion in a substrate-dependent manner^{13,14}. The C-domain active site, but not the N domain, in ACE is strongly activated by chloride ion¹⁵. The structure of tACE revealed the location of two buried chloride ions separated by 20.3 Å (Fig. 1a). The first (Cl1, 20.7 Å away from the zinc ion) is bound to four ligands—Arg 489 (NH1), Arg 186 (NE), Trp 485 (NE1) and water—and is surrounded by a hydrophobic shell of four tryptophans. The second (Cl2, 10.4 Å away from the zinc ion) is bound to Arg 522 (NE, 3.1 Å), in agreement with a previous report indicating that Arg 1098 (the analogous Arg residue in the C domain of somatic ACE) is critical for the chloride dependence of ACE activity¹⁶. Tyr 224 and a water molecule are the other two Cl2 ligands. Thus, binding of each chloride in tACE involves ligands from both subdomains (Fig. 1c). It has been proposed that the chloride ion might interact with the substrate as well as the

enzyme¹⁶. However, the location of both buried chloride ions outside the active site makes such an interaction unlikely. The molecular mechanism of chloride activation is not readily apparent from the structure; however, the primary ligand for Cl2, Arg 522, lies on the same helix ($\alpha 17$) as two residues (Tyr 520 and Tyr 523) that interact with lisinopril and, presumably, substrate as well (Figs 1c and 2a).

ACE belongs to the M2 family of zinc-binding metalloproteases, within the MA clan. Apart from the HEXXH metal-coordinating motif, there is little sequence homology between ACE and other members of the MA clan. Structural comparison of tACE with other protein structures using the DALI server¹⁷ revealed significant homology with neurolysin², a protein involved in neurotensin metabolism (Fig. 3a, b), and a carboxypeptidase from the hyperthermophilic archaeon *P. furiosus*³. Neurolysin is a member of the M3 family of oligopeptidases (a member of the MA clan), and *P. furiosus* carboxypeptidase is a member of the M32 family of carboxypeptidases. Similar to ACE, both belong to the family of metalloproteases bearing the HEXXH active-site motif¹⁸ and consist of an abundance of α -helices with very little β -structure. The two proteins exhibit little amino-acid sequence similarity with tACE, yet when the two structures are optimally superimposed, there is a noticeable match with an r.m.s. deviation of 3.4 Å for 361 C α atoms and 3.11 Å for 399 C α atoms against neurolysin and *P. furiosus* carboxypeptidase, respectively. The core structures for

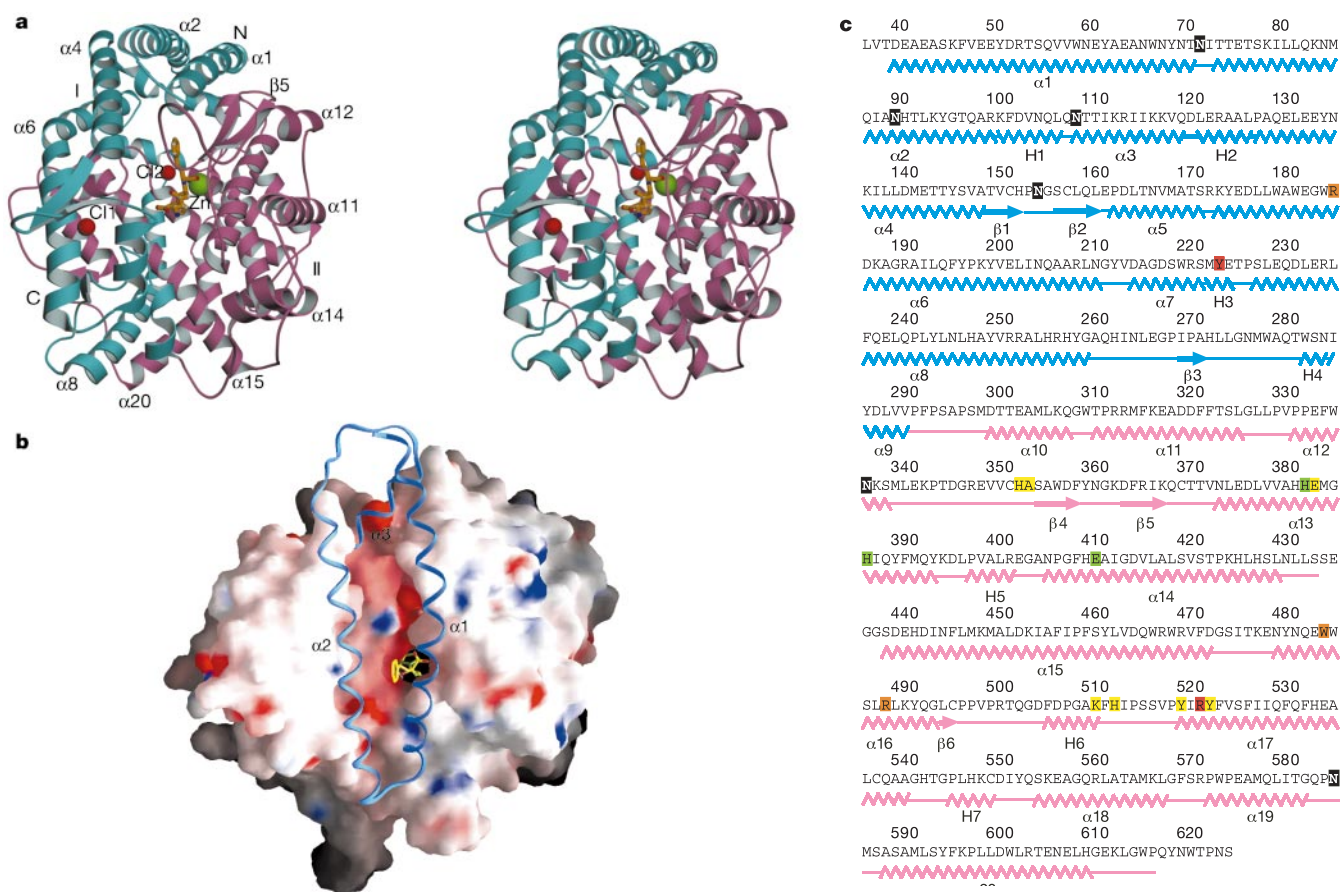


Figure 1 Overview of tACE structure. **a**, Stereo view of the ribbon representation of the molecule looking down on the active site. The molecule can be divided into two portions, as subdomains I and II (cyan and pink, respectively). The active-site zinc ion and the lisinopril molecule are shown in green and yellow, respectively. The two chloride ions are shown as red spheres. **b**, Molecular surface representation (negative and positive potentials in red and blue, respectively) showing the active-site groove. The view is at 90° (towards the observer) to **a**. For clarity, the molecular surface has been sliced. The buried

lisinopril molecule is shown in yellow. Helices $\alpha 1$, $\alpha 2$ and $\alpha 3$ forming the lid are shown. **c**, The structure–sequence relationship in tACE¹⁰. The secondary structure elements (subdomain I in cyan; subdomain II in pink) follow the same colour code as in **a**. α , α -helices; β , β -strands; H, 3_{10} helices. The important residues that are involved in binding are indicated as follows: zinc ligands, green boxes; chloride-binding residues, orange (Cl1) and red (Cl2) boxes; lisinopril-binding residues, yellow boxes; and glycosylation sites, black boxes.

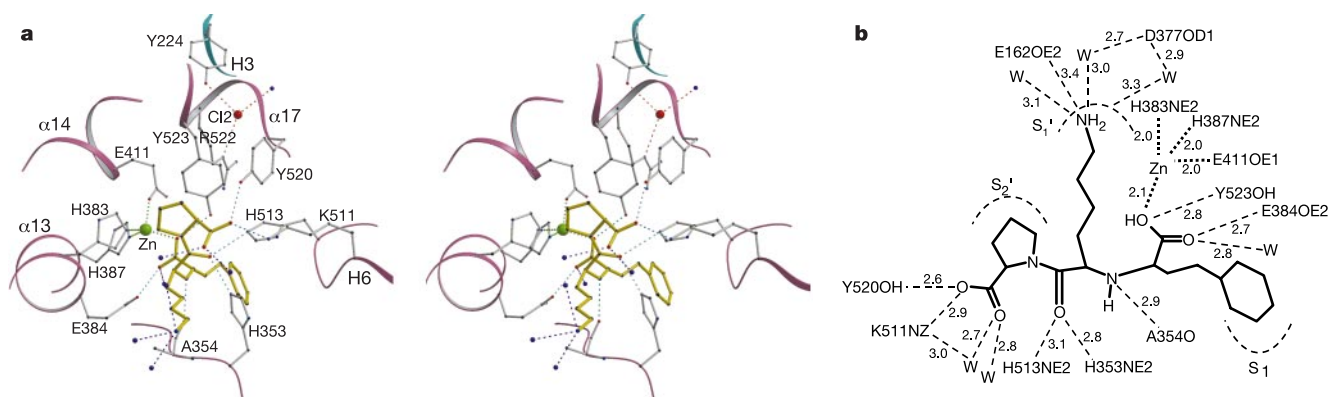


Figure 2 Details of the active site. **a**, Binding of lisinopril to tACE (stereo representation). Selected residues are shown in a 'ball-and-stick' representation with zinc atoms in green, chloride ions in red, water molecules in purple, and lisinopril (inhibitor) in yellow. Important secondary structure elements are marked. The lisinopril, Cl⁻, water hydrogen

bonds and zinc coordination are shown in cyan, red, purple and green dotted lines, respectively. **b**, Schematic view of lisinopril binding with distances marked in Å. The different binding subsites are labelled.

these three proteins are similar, with significant differences in loops on the outer surface. The marked similarity also extends to the active-site region in neurolysin and *P. furiosus* carboxypeptidase, which consists of a deep narrow channel (about 20% larger accessible surface area than in tACE) that divides the molecule into two halves, and which only allows access to small peptide substrates. Notably, ACE also acts on neurotensin among its wide range of substrates¹⁹. Structural comparison of tACE with thermolysin (M4 family) revealed that there was significant, but far less marked, structural similarity between them (r.m.s. deviation of 3.2 Å for 170 C α atoms).

The structure of tACE bound to the potent inhibitor lisinopril^{6,20,21} ($K_i = 2.7 \times 10^{-10}$ M) shows that the inhibitor binds in a highly ordered (overall *B*-factor of 15.26 Å²), extended conformation, with the phenyl group stretching towards the lid (Fig. 1a, b) and the lysine side chain parallel to the α 13 helix containing the HEXXH motif (Fig. 2a). The lisinopril molecule is buried about 10 Å inside the groove (as viewed in Fig. 1a), in agreement with the previous studies on biotinylated derivatives of lisinopril with ACE²². No significant rearrangement of active-site residues was observed on complex formation and the chloride ions are bound as in the native structure. The carboxylate of lisinopril is well positioned to bind to the active-site zinc atom^{20,21}, and provides one coordinating ligand (provided by an acetate ion in the native structure). The other three ligands (two histidines and a glutamic acid) are the same in both structures. The second oxygen atom of this carboxylate is 2.6 Å away from the zinc atom and makes an

H-bond interaction with Glu 384 (OE2 atom, which appears to be protonated, as the complex was crystallized at about pH 4.7) of the HEXXH motif (Fig. 2a, b). The S₁ phenylpropyl group makes van der Waals interactions with Val 518, and the lysyl amine forms a weak H-bond with Glu 162 (OE2 atom, 3.4 Å) at the S₁' subsite of tACE. Surprisingly, the C-terminal carboxylate, which was thought to interact with a positively charged arginine residue (based on chemical modification experiments²³), instead binds to a lysine (Lys 511) as well as to Tyr 520.

We have presented here a three-dimensional structure of tACE and the tACE–lisinopril complex. Our study provides a detailed picture of the active site, which could be used for the development of new, highly selective ACE inhibitors targeted to the C domain by structure-based, rational drug design. This approach should produce a new generation of ACE inhibitors with the potential for greater efficacy, fewer side effects and new treatment indications (for example, polycythaemia). Furthermore, the similarity of tACE with neurolysin and *P. furiosus* carboxypeptidase has shown a surprising structural conservation between these metalloproteases. □

Methods

Purification, crystallization and data collection

Details of the production of recombinant tACE (tACE Δ 36NJ) and growth of crystals (by vapour diffusion) will be reported elsewhere (our own unpublished results). Briefly, 2 μ l of the protein solution was mixed in 10 mM HEPES and 0.1% phenylmethylsulphonyl fluoride, with an equal volume of a reservoir solution containing 15% PEG 4000, 50 mM sodium acetate trihydrate, pH 4.7, and 10 μ M ZnSO₄·7H₂O. tACE–lisinopril complex crystals were grown at 1:1.2 (protein/inhibitor) ratio of the inhibitor. Both native and complex crystals belonged to the P2₁2₁2₁ space group (grown at 16 °C, cell dimensions: *a* = 56.47, *b* = 84.90, *c* = 133.99 Å) with one molecule in the asymmetric unit. All X-ray data collection was performed at 100K either at the European Synchrotron Radiation Facility (ESRF) or at the Synchrotron Radiation Source (SRS). Both native and lisinopril complex data sets were collected to 2.0 Å (*R*_{merge} (native), 8.1% (last shell, 43.2%) with overall completeness 98.8% (2.07–2.0 Å, 95.7%); *R*_{merge} (tACE–lisinopril) 5.8% (last shell, 23.8%) with overall completeness 95.9% (2.08–2.01 Å, 72.1%)). For structure solution, multiple anomalous dispersion (MAD) data were collected at the peak of the zinc-K edge (1.2825 Å), inflection (1.2832 Å) and a remote wavelength (0.9537 Å), and at 1.7712 Å, the closest possible wavelength at the sulphur K-edge. Out of 32 soaks with different heavy atoms, only 3 were found to be useful. These derivatives were prepared by soaking the crystals (approximately 10–60 min, 1–5 mM concentration) in the presence of K₂PtCl₄, K₂PdCl₄ and OsCl₃, respectively. All data were processed using the program HKL2000 (ref. 24).

Structure determination and refinement

The crystal structure of the ACE–lisinopril complex was solved using MAD and Multiple Isomorphous Replacement with Anomalous Scattering (MIRAS) procedures. The position of the zinc atom was identified using anomalous difference Patterson maps calculated using diffraction data at peak wavelength. The MAD phases that we obtained were not very strong, and therefore additional phase information was obtained using MIRAS procedures with platinum (K₂PtCl₄), palladium (K₂PdCl₄) and osmium (OsCl₃)

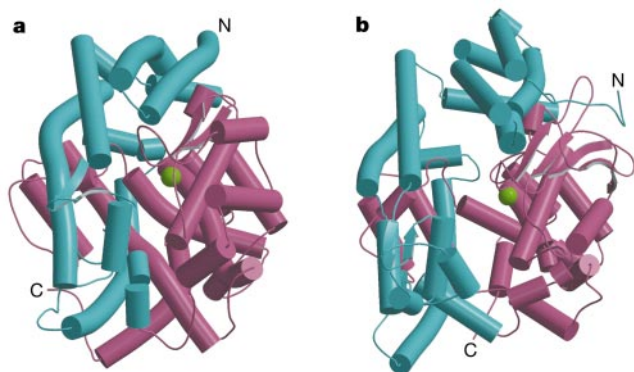


Figure 3 Comparison of tACE (**a**) and neurolysin (**b**) folds. The same view as in Fig. 1a is retained.

heavy-atom derivatives. The zinc site was used to obtain the starting phases for each derivative. Double difference Fourier maps calculated using the CCP4 program²⁵ gave the first major heavy-atom site, and the phases from the combined zinc and heavy-atom sites were used to locate additional major/minor sites for each derivative. All heavy-atom sites and the zinc site were refined to 2.8 Å resolution using both MLPHARE²⁵ and SHARP²⁶. The overall figure of merit from SHARP was 0.38 (at 2.8 Å) and was improved to 0.89 (at 2.0 Å) by iterative solvent flattening, phase combination, and phase extension with the program SOLOMON²⁷. We carried out model building and refinement using the programs O and CNS²⁸, respectively. During the final stages of refinement, water molecules, zinc ion, acetate ion and the inhibitor molecule were inserted in the respective structures. Both the native and inhibitor complex structures were refined at 2.0 Å (see Supplementary Information) with 94% of the residues in the maximum allowed region and none in the disallowed region of the Ramachandran map. The final structures of ACE and the ACE-lisinopril complex had an R_{free} value of 22.08% and 21.88%, and a final R_{cryst} value of 18.29% and 18.14%, respectively. Both models have r.m.s. deviations in bond length of 0.005 Å and in bond angles of 1.2°. Diagrams were computed using MOLSCRIPT²⁵, RASTER3D²⁹ (Figs 2a and 3a, b), POVRAY (Fig. 1a; see <http://www.povray.org>) and GRASP³⁰ (Fig. 1b).

Received 21 October; accepted 17 December 2002; doi:10.1038/nature01370.
Published online 19 January 2003.

- Cushman, D. W., Cheung, H. S., Sabo, E. F. & Ondetti, M. A. Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. *Biochemistry* **16**, 5484–5491 (1977).
- Brown, C. K. *et al.* Structure of neurolysin reveals a deep channel that limits substrate access. *Proc. Natl Acad. Sci. USA* **98**, 3127–3132 (2001).
- Arndt, J. W. *et al.* Crystal structure of a novel carboxypeptidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Structure* **10**, 215–224 (2002).
- Soubrier, F. *et al.* Two putative active centers in human angiotensin I-converting enzyme revealed by molecular cloning. *Proc. Natl Acad. Sci. USA* **85**, 9386–9390 (1988).
- Williams, T. A., Corvol, P. & Soubrier, F. Identification of two active site residues in human angiotensin I-converting enzyme. *J. Biol. Chem.* **269**, 29430–29434 (1994).
- Wei, L., Clauser, E., Alhenc-Gelas, F. & Corvol, P. The two homologous domains of human angiotensin I-converting enzyme interact differently with competitive inhibitors. *J. Biol. Chem.* **267**, 13398–13405 (1992).
- Junot, C. *et al.* RXP 407, a selective inhibitor of the N-domain of angiotensin I-converting enzyme, blocks *in vivo* the degradation of hemoregulatory peptide acetyl-Ser-Asp-Lys-Pro with no effect on angiotensin I hydrolysis. *J. Pharmacol. Exp. Ther.* **297**, 606–611 (2001).
- Esther, C. R. *et al.* The critical role of tissue angiotensin-converting enzyme as revealed by gene targeting in mice. *J. Clin. Invest.* **99**, 2375–2385 (1997).
- Ehlers, M. R. W., Fox, E. A., Strydom, D. J. & Riordan, J. F. Molecular cloning of human testicular angiotensin-converting enzyme: the testis isozyme is identical to the C-terminal half of endothelial angiotensin-converting enzyme. *Proc. Natl Acad. Sci. USA* **86**, 7741–7745 (1989).
- Yu, X. C. *et al.* Identification of N-linked glycosylation sites in human testis angiotensin-1 converting enzyme and expression of an active deglycosylated form. *J. Biol. Chem.* **272**, 3511–3519 (1997).
- Chubb, A. J., Schwager, S. L. U., Woodman, Z. L., Ehlers, M. R. W., Sturrock, E. D. *Biochem. Biophys. Res. Commun.* **297**, 1225–1230 (2002).
- Ehlers, M. R. W. & Riordan, J. F. Angiotensin-converting enzyme: zinc- and inhibitor-binding stoichiometries of the somatic and testis isozymes. *Biochemistry* **30**, 7118–7126 (1991).
- Bünning, P. & Riordan, J. F. Activation of angiotensin converting enzyme by monovalent anions. *Biochemistry* **22**, 110–116 (1983).
- Shapiro, R., Holmquist, B. & Riordan, J. F. Anion activation of angiotensin converting enzyme: dependence on nature of substrate. *Biochemistry* **22**, 3850–3857 (1983).
- Jaspard, E., Wei, L. & Alhenc-Gelas, F. Differences in the properties and enzymatic specificities of the

- two active sites of angiotensin I-converting enzyme (Kininase II). *J. Biol. Chem.* **268**, 9496–9503 (1993).
- Liu, X., Fernandez, M., Wouters, M. A., Heyberger, S. & Husain, A. Arg-1098 is critical for the chloride dependence of human angiotensin-1 converting enzyme C-domain catalytic activity. *J. Biol. Chem.* **276**, 33518–33525 (2001).
- Holm, L. & Sander, C. Protein folds and families: sequence and structure alignments. *Nucleic Acids Res.* **27**, 244–247 (1999).
- Corvol, P. & Williams, T. A. in *Handbook of Proteolytic Enzymes* (eds Barrett, A. J., Rawlings, N. D. & Woessner, J. F.) 1066–1076 (Academic, London, 1998).
- Skidgel, R. A., Engelbrecht, S., Johnson, A. R. & Erdos, E. G. Hydrolysis of substance P and neurotensin by converting enzyme and neutral endopeptidase. *Peptides* **5**, 769–776 (1984).
- Patchett, A. A. *et al.* A new class of angiotensin-converting enzyme inhibitors. *Nature* **288**, 280–283 (1980).
- Patchett, A. A. & Cordes, E. H. The design and properties of N-carboxyalkyldipeptide inhibitors of angiotensin converting enzyme. *Adv. Enzymol.* **57**, 1–84 (1985).
- Bernstein, K. E., Welsh, S. L. & Inman, J. K. A deeply recessed active site in angiotensin converting enzyme is indicated from the binding characteristic of biotin-spacer-inhibitor reagents. *Biochem. Biophys. Res. Commun.* **167**, 310–317 (1990).
- Bunning, P., Holmquist, B. & Riordan, J. F. Functional residues at the active site of angiotensin converting enzyme. *Biochem. Biophys. Res. Commun.* **83**, 1442–1449 (1978).
- Otwinowski, M. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Collaborative computational project Number 4 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- De La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameters refinement in the MIR and MAD methods. *Methods Enzymol.* **276**, 472–494 (1997).
- Abrahams, J. P. & Leslie, A. G. W. Methods used in structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 110–119 (1996).
- Brünger, A. T. *et al.* Crystallography & N. M. R. system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Merritt, E. A. & Bacon, D. J. Raster 3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrogen atoms. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This work was supported by a Wellcome Trust grant to K.R.A., a Collaborative Research Initiative Grant to E.D.S. and K.R.A., and a National Research Foundation Grant to E.D.S. We thank M. Walsh, H. Belrahli and A. Thompson at ESRF, and L. Duke, J. Nicholson, M. McDonald, P. Rizkallah and M. Papiz at SRS for their help during X-ray data collection. We also thank R. Shapiro, J. Riordan and M. Ehlers for constructive criticism of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.R.A. (e-mail: bskkra@bath.ac.uk) or E.D.S. (e-mail: sturrock@curie.uct.ac.za). The X-ray coordinates of tACE and the tACE-lisinopril complex have been deposited in the Protein Data Bank under entry codes 1O8A and 1O86, respectively.

Proteomics paraphernalia

Synthesis, analysis and screening.

Assays-on-Demand

Applied Biosystems

www.europe.appliedbiosystems.com

Ready-to-use genetic screening assays

Applied Biosystems offers 77,000 assays for SNP genotyping and 4,400 assays for quantitative gene expression analysis. All assays are optimized for use on the ABI PRISM 7900Ht, 7700 and 7000 sequence detection systems. The assays are functionally validated and designed specifically for genetic screening.

QIAamp MinElute

Qiagen

www.qiagen.com

Pure and simple

MinElute kits can be used to isolate viral RNA and DNA simultaneously from a broad range of RNA and DNA viruses. Elution in small volumes yields highly concentrated nucleic acids in less than one hour. Nucleic acids are ready for immediate use in amplification reactions such as PCR and RT-PCR, or for storage at -20°C . The kits are suitable for use with plasma, serum and other cell-free body fluids.

ProteinDetector

KPL

www.kpl.com

Alkaline phosphatase ELISA kits

These kits have been designed to offer a convenient starting point for the development of microwell ELISA procedures for antigen or antibody detection. They include pre-tested reagents, including blocking, coating and wash solutions, alkaline phosphatase-labelled secondary antibodies, BluePhos substrate and a detailed instruction manual.

Q-Mate

Qbiogene

www.qbiogene.com

Use it on and off

The effects of toxic proteins can be studied *in vitro* and *in vivo* with this system. Q-Mate uses molecular expression control, allowing researchers to turn expression of the toxic protein off, infect the target cell with the gene of interest, and then turn gene expression back on to study the effects and function of the protein that is expressed. It is based on the regulatory gene *CymR*, which controls the transformation of *p*-cymene to cumate and the further degradation of the cumate molecule in *Pseudomonas putida* F1. A recombinant adenovirus is used to carry the gene of interest and infect the target cell.



PolyPlex: oligonucleotides, and quick.

PolyPlex

GeneMachines

www.genemachines.com

Oligonucleotide synthesizer

The newly redesigned PolyPlex 96-well oligonucleotide synthesizer generates a full plate of 20mers in less than three hours. It can synthesize up to one micromole scale using cartridges and accommodating a variety of chemistries, with eight amidite reagent positions. It is compatible with a wide range of reagent bottle sizes.

GenePix 4100A/ Pro 4.1

Axon Instruments

www.axon.com

Personal scanner

Axon describes its GenePix 4100A microarray scanner as a compact, easy-to-use device with a high level of performance. It is controlled using GenePix Pro 4.1, a software program for microarray analysis. The program is compatible with all GenePix scanners and can be linked to Acuity, the company's microarray database, data mining and visualization software. It is also available as a stand-alone package for analysis of 16-bit TIFF images from third-party scanners.

GenVision

DNASTAR

www.dnastar.com

Put words into pictures

This data visualization software creates Postscript images from annotation data stored in text files. Both linear and circular genome maps can be drawn. GenVision is available as a plug-in for Adobe Illustrator for Macs and Windows, or in a stand-alone version for Windows. The plug-in provides dialogs and tools to add GenVision functionality to Illustrator. The stand-alone version provides all of the GenVision image creation functions in a less complex environment.

PicoPure

Arcturus

www.arctur.com

Proteomic studies on pure cell populations

PicoPure is a general-purpose chaotropic buffer validated for protein extraction from laser capture microdissected cells. The kits contain 10 ml of buffer and complete protocols for protein extraction from samples on CapSure LCM caps.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

 Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

 Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

Visions of Europe's future

What would it take to reverse the brain drain that has drawn so many Europeans to work in the United States? Here's a possible scenario. First, the budget for the US National Institutes of Health abruptly ends its five-year doubling trend, leaving research funding more or less flat. The National Science Foundation continues to make modest budgetary gains. But the main result is that few new grants are funded — unless they are related to biodefence.

Meanwhile, European countries raise their commitment to research funding. Foundations and societies launch schemes to lure top talent back home. And governments, businesses and universities combine to create institutions that are funded by, but relatively free from the control of, all three.

Far-fetched? Not really. The US factors are close to reality. And the European components are starting to materialize. In this week's instalment of Movers (moved from the back of the journal to a new slot at the front of the *Naturejobs* section), hints of each factor have contributed to some examples of repatriation.

Signs of an increased commitment to science in Britain, a new chemistry facility by Oxford and start-up money from the Royal Society and Wolfson Foundation lured Hagan Bayley from Texas A&M University back to the University of Oxford, where he studied as an undergraduate.

But for a true reversal of fortunes, a few more factors must be addressed. European countries need to raise their overall levels of research and development. They also need to offer packages that allow investigators to fund postdocs, graduate students and technicians. And they should promote autonomy — whether it be freedom to work with industry, or to avoid it. Without those things, repatriation will remain the exception rather than the rule.

Paul Smaglik

Naturejobs editor



Contents

MOVERS

Oxford professor comes home;
physicist heads for Harvard;
linguistic challenges at Aventis;
and more p558

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Transitions

X Wendy Freedman will next month succeed **Augustus Oemler** as director of the Carnegie Observatories in Pasadena, California. Freedman first joined the observatories as a postdoc in 1984.

X TransForm Pharmaceuticals, a biotech company in Lexington, Massachusetts, this month took on **Duncan Higgons** as chief commercial officer. Higgons joins the company from drug-delivery firm Alkermes in Cambridge, Massachusetts, where he was senior vice-president of business development and marketing.

X BioXell, a biopharmaceutical company based in Milan, Italy, this month appointed **Alex Martin** as chief business officer. Martin comes to BioXell from Novartis Pharma in Basel, Switzerland, where he was vice-president for global business development and licensing.

X Robert Liptak, the principal and chief financial officer at global investment management firm MPM Capital's Boston office, was last month promoted to general partner and chief financial officer. The move follows the appointment in November of **Markus Hosang**, formerly of Hoffmann-La Roche in Basel, Switzerland, as a new venture partner in MPM's Munich office.

X Paul Lunn this month joins Natestech Pharmaceutical in Bothell, Washington, as patent counsel. He was most recently principal patent attorney at ZymoGenetics, a Seattle-based biopharmaceutical company. **Gordon Brandt** joined Natestech last month as executive vice-president of science and clinical development.

CHEMISTRY

When **Hagan Bayley** takes up his position as chair of chemical biology at the University of Oxford in October, it will mark a homecoming. Bayley, head of the medical biochemistry and genetics department at Texas A&M University, graduated from Balliol College Oxford before heading to Harvard, where he gained his PhD in 1979. Since then he has worked mainly in the United States, including posts at the Massachusetts Institute of Technology and Columbia University.

Funding from the Royal Society and the Wolfson Foundation, designed to encourage 'brain gain' to Britain, helped to attract him back, as did a £60-million (US\$96-million) chemistry research laboratory at Oxford, due to be completed this summer. The building will help to bring together physical and organic chemists, who, Bayley says, have been 'balkanized' in separate buildings in the past.

PHARMACEUTICALS

Although **Ruth Roberts** has been at Aventis's Paris office since last summer, she felt she'd officially arrived this month, when she held her first departmental meeting in French. Roberts moved after 12 years with Syngenta's Central Toxicology Laboratory in the United Kingdom, where she was head of cancer biology research. She went to Paris as director of toxicology for Aventis Drug Safety partly to immerse herself in another culture, she says, and partly to expand her expertise in regulatory toxicology.

With her background in the molecular mechanisms of adverse responses to drugs, Roberts now enjoys bridging research and regulatory affairs. "I can anticipate research needs for safety assessments," she says. As for the language, it's coming along. After her first French-speaking meeting, she checked in English to see whether people had done or planned to do various things. "Tenses can be very important," she jokes.

UNIVERSITY ADMINISTRATION

After 30 years away, **Jack Dixon** returns to the University of California next month as Dean for Scientific Affairs, Health Sciences, at San Diego (UCSD). Dixon did his first degree at Los Angeles, his PhD at Santa Barbara and a postdoc at San Diego.

"The University of California system has been awfully good to me in terms of providing training," Dixon says. "It's nice to be able to give something back." He moved in 1973 to Purdue University and in 1991 to the University of Michigan, where he was most recently director of the Life Sciences Institute.

His biggest challenge at UCSD may be to



Hagan Bayley



Ruth Roberts



Jack Dixon



Federico Capasso



Kristian Helin

manage growth in the middle of a state budget crisis (see *Nature* **421**, 202; 2003). For instance, he admits it is unlikely that the UCSD campus will double in size, as planned, during the next ten years. But he sees room for optimism, including a new school of pharmacy and the fact that many medical-faculty members will retire in the next ten years, allowing him to bring in new talent. "All the gloomy news I take seriously, but I don't view it as something that will last very long," he says.

Dixon, an expert on the role played by phosphatase enzymes in cellular responses to molecular signals, will maintain a lab at UCSD and will also hold professorships in pharmacology and in cellular and molecular medicine.

PHYSICS

After 26 years at Bell Laboratories in Murray Hill, New Jersey, most recently as vice-president of physical research, **Federico Capasso** this month joined Harvard University as Gordon McKay professor of applied physics and Vinton Hayes senior research fellow in electrical engineering. Known for his pioneering research on semiconductor nanostructures, Capasso helped to invent the quantum cascade laser, which has applications in many disciplines including planetary science, environmental monitoring, atmospheric chemistry and medical diagnostics.

He is pleased by Harvard's increasing commitment to interdisciplinary research. "Nourishing those interfaces has always appealed to me," Capasso says. "It's extremely attractive for me to break down barriers and establish collaborations with lots of professors."

BIOLOGY

Kristian Helin, a Dane who since 1995 has worked in Milan at the European Institute of Oncology, is leaving Italy to direct Denmark's new flagship Biotech Research and Innovation Centre (BRIC). "This is a position where I hope I can have a very big impact on research going on in Denmark," he says. BRIC is run by an independent board, associated with the University of Copenhagen but with industry representation too. "It gives you the possibility to move faster than you normally could in a university and have good contact with industry," says Helin.

The 3,000-square-metre facility is due to open in 2005; until then, BRIC will lease space from the university. Helin aims to start recruiting scientists for 12–15 research groups this year, with at least 30% of them coming from outside Denmark.

CONTACTING US AT MOVERS

Please send any information on appointments or job moves to:

✉ naturejobseditor@naturedc.com